GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: May 7, 2003, 09:27:25 ; Search time 35 Seconds
(without alignments)
87.565 Million cell updates/sec

Title: US-09-674-973A-17
Perfect score: 106
Sequence: 1 SLVRLSSCVPVALMSAMTTSSSQ 23

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	106	100.0	23	21	AAY65700	TGF beta RII mutan
2	106	100.0	34	17	AAW05380	Fragment of VACO45
3	106	100.0	34	21	AAY54017	Peptide which is n
4	106	100.0	34	21	AAY65696	TGF beta RII mutan
5	106	100.0	34	22	AAB82997	Truncated TGF-beta
6	106	100.0	34	23	ABB80865	Type II TGFbeta re
7	106	100.0	44	21	AAY65697	TGF beta RII mutan
8	79	74.5	18	21	AAY66121	Frameshift mutated
9	64	60.4	26	21	AAY54038	Peptide used to pr
10	57	53.8	22	21	AAY65701	TGF beta RII mutan
11	53	50.0	11	21	AAY54019	Peptide which is c
12	53	50.0	23	21	AAY65699	TGF beta RII mutan
13	49	46.2	10	21	AAY54022	Peptide which is c
14	49	46.2	113	22	AAG73473	Human gene 17-enco
15	46	43.4	138	22	AAU67640	Propionibacterium
16	46	43.4	194	22	AAU55878	Propionibacterium
17	46	43.4	225	22	AAB83360	NOV4 protein seque
18	45	42.5	9	21	AAY54021	Peptide which is c
19	45	42.5	9	21	AAY54037	Peptide which is c
20	45	42.5	9	21	AAY66122	Frameshift mutated
21	45	42.5	9	21	AAY66123	Frameshift mutated
22	45	42.5	9	21	AAY66124	Frameshift mutated
23	45	42.5	9	21	AAY66125	Frameshift mutated
24	45	42.5	9	21	AAY66126	Frameshift mutated
25	45	42.5	9	21	AAY66129	Frameshift mutated
26	45	42.5	697	20	AAY31753	Candida cloacae fa
27	44.5	42.0	134	22	AAG72318	Human olfactory re
28	44	41.5	9	21	AAY54035	Peptide which is c
29	44	41.5	9	21	AAY66127	Frameshift mutated
30	44	41.5	34	22	ABB42070	Peptide #9576 enco
31	44	41.5	34	22	ABB25671	Protein #7670 enco
32	44	41.5	34	22	AAM62951	Human brain expres
33	44	41.5	34	22	AAM75763	Human bone marrow
34	44	41.5	34	22	AAM20685	Peptide #7119 enco
35	44	41.5	34	22	AAM35871	Peptide #9908 enco
36	44	41.5	34	23	ABG45243	Human peptide enco
37	44	41.5	745	22	AAB85499	Human protein kina
38	43.5	41.0	94	18	AAW22506	Sugar beet antimic
39	43	40.6	43	21	AAB12327	Human secreted pro
40	43	40.6	53	22	AAU58045	Propionibacterium
41	43	40.6	74	22	AAU61820	Propionibacterium
42	43	40.6	119	23	ABP08308	Human ORFX protein
43	43	40.6	128	23	ABB89867	Human polypeptide
44	43	40.6	157	22	AAU48217	Propionibacterium
45	43	40.6	272	23	ABB53632	Lactococcus lactis

ALIGNMENTS

RESULT 1
AAY65700
ID AAY65700 standard; Peptide; 23 AA.
XX
AC AAY65700;
XX
DT 10-FEB-2000 (first entry)
XX
DE TGF beta RII mutant peptide 5.
XX
KW Human; frameshift mutant; T cell response; tumour; treatment; cancer;
KW mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9958552-A2.
XX
PD 18-NOV-1999.
XX
PF 03-MAY-1999; 99WO-NO00143.
XX
PR 08-MAY-1998; 98NO-0002097.
XX
PA (NHYD) NORSK HYDRO AS.
XX
PI Gaudernack G, Eriksen JA, Moller M, Gjertsen MK, Saeterdal I;
XX
DR WPI; 2000-039064/03.
XX
PT New peptides derived from genes with frameshift mutations, used to
PT develop products for the treatment and prophylaxis of cancers -
XX

PS Claim 12; Page 20; 166pp; English.
XX
CC Peptides AAY65684-Y66142 are fragments of mutant proteins arising from a
CC frameshift mutation in a gene from a cancer cell. The peptides are
CC characterised in that they:
CC (i) are at least 8 amino acids long and a fragment of a mutant protein
CC arising from a frameshift mutation in a gene of a cancer cell;
CC (ii) consist of at least one amino acid of the mutant part of a protein
CC sequence encoded by the gene;
CC (iii) comprise 0-10 amino acid from the carboxyl terminus of the normal
CC part of the protein sequence preceding the amino terminus of the mutant
CC sequence and may further extend to the carboxyl terminus of the mutant
CC part of the protein as determined by a new stop codon generated by the
CC frameshift mutation; and
CC (iv) induce, either in their full lengths or after processing by an
CC antigen presenting cell (APC), T cell responses.
CC The genes that the peptides are derived from, are characterised as
CC susceptible to frameshift mutation by having a mono nucleoside base
CC repeat sequence of at least 5 residues, or a di-nucleoside base repeat
CC sequence of at least 4 di-nucleoside base units. The peptides are
CC created by the addition or deletion of 1 or 2 nucleoside base residues
CC from the repeat sequence. The novel peptides can elicit T cell responses
CC and toxicity against tumours and cancer cells carrying genes with
CC frameshift mutations. The novel peptides and DNA sequences can be used
CC for the preparation of a composition for the treatment or prophylaxis of
CC cancer.
XX
SQ Sequence 23 AA;

Query Match 100.0%; Score 106; DB 21; Length 23;
Best Local Similarity 100.0%; Pred. No. 2.7e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLVRLSSCVPVALMSAMTTSSSQ 23
||||||||||||||||||||||||
Db 1 SLVRLSSCVPVALMSAMTTSSSQ 23

RESULT 2
AAW05380
ID AAW05380 standard; peptide; 34 AA.
XX
AC AAW05380;
XX
DT 04-JUN-1997 (first entry)
XX
DE Fragment of VACO457 RII mutant.
XX
KW Type I transforming growth factor beta receptor gene; epithelial cell;
KW tumour development; cancer; non-functional mutant; precancerous lesion;
KW growth regulatory gene; type II receptor; serine/threonine receptor;
KW tumour tissue; colonic cancer; endometrial cancer; ovarian cancer;
KW gastric cancer; TGFbeta receptor gene; pancreatic cancer.
XX
OS Synthetic.
XX
PN WO9631605-A1.
XX
PD 10-OCT-1996.
XX
PF 05-APR-1996; 96WO-US04727.
XX
PR 22-MAY-1995; 95US-0445520.
PR 07-APR-1995; 95US-0417867.
XX
PA (MEDI-) MEDICAL COLLEGE OHIO.
PA (UYCA-) UNIV CASE WESTERN.
XX
PI Brattain MG, Markowitz SD, Willson JKV;
XX
DR WPI; 1996-465028/46.
XX
PT Cancer diagnosis and therapy - based on mutation(s) in type II
PT transforming growth factor beta receptor
XX
PS Disclosure; Page 30; 70pp; English.
XX
CC This sequence represents a fragment of the type II transforming growth
CC factor beta (TGFbeta) receptor gene mutant VACO457. TGFbeta inhibits the
CC growth of multiple epithelial cell types, and loss of this negative
CC regulation is thought to contribute to tumour development. TGFbeta also
CC inhibits the growth of certain cancer cell lines. This sequence can be
CC detected by a method of the invention. The method of the invention is for
CC aiding cancer diagnosis or prognosis. The method comprises detecting
CC expression of a mutant form of type II TGFbeta receptor (mutant RII) by
CC cells of a patient or the absence of wild-type RII in tumour cells.
CC Another method comprises detecting a non-functional mutant form of a
CC growth regulatory gene which encodes a type II receptor which is a member
CC of a family of serine/threonine receptors that bind members of a family
CC of TGFbeta-like factors. Alternatively, the method comprises detecting a
CC mutant growth regulatory gene which contains repetitive DNA sequence
CC motifs in the wild-type coding region, where the presence of the
CC non-functional mutant form is indicative of tumour tissue or precancerous
CC lesions. The methods can be used for diagnosis or treatment of colonic,
CC endometrial, ovarian, gastric or pancreatic cancer or other malignancies.
XX
SQ Sequence 34 AA;

Query Match 100.0%; Score 106; DB 17; Length 34;
Best Local Similarity 100.0%; Pred. No. 4.3e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLVRLSSCVPVALMSAMTTSSSQ 23
||||||||||||||||||||||||
Db 1 SLVRLSSCVPVALMSAMTTSSSQ 23

34.

RESULT 3
AAY54017
ID AAY54017 standard; peptide; 34 AA.
XX
AC AAY54017;
XX
DT 27-MAR-2000 (first entry)
XX
DE Peptide which is not a part of MHCI glycoprotein binding peptides.
XX
KW Class I major histocompatibility glycoprotein complex; MHCI;
KW mutant RII receptor; transforming growth factor-beta; TGF-beta;
KW tumour; vaccine; gastric cancer; colon cancer; gene therapy.
XX
OS Homo sapiens.
XX
PN FR2779432-A1.
XX
PD 10-DEC-1999.
XX
PF 08-JUN-1998; 98FR-0007322.
XX
PR 08-JUN-1998; 98FR-0007322.
XX
PA (TRGE) TRANSGENE SA.
XX
DR WPI; 2000-074958/07.
XX
PT New nucleic acid sequences, useful for production of medicament for
PT diagnosing, preventing and/or treating gastric or colon cancers -
XX
PS Claim 1; Page 19; 41pp; French.
XX
CC The specification describes peptides which are capable of fixing
CC themselves on at least one class I major histocompatibility
CC glycoprotein complex (MHCI), and which do not comprise the present
CC sequence. The peptides are derived from a mutant RII receptor of
CC transforming growth factor-beta (TGF-beta). The presence of the
CC mutant receptor leads to inactivation of TGF-beta, and contributes

CC to the development of tumours. Especially, the mutation comprises
CC the addition or deletion of an adenine between positions 709-718.
CC The peptides, or nucleic acids encoding them, are useful for the
CC production of a medicament (either preventative, therapeutic or
CC as a vaccine) for treating gastric cancers or cancers of the colon
CC by gene therapy or the peptide may be used as a diagnostic,
CC prophylactic and/or therapeutic composition for the detection,
CC prevention or treatment of gastric or colon cancers.
XX
SQ Sequence 34 AA;

Query Match 100.0%; Score 106; DB 21; Length 34;
Best Local Similarity 100.0%; Pred. No. 4.3e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLVRLSSCVPVALMSAMTTSSSQ 23
||||||||||||||||||||||||
Db 1 SLVRLSSCVPVALMSAMTTSSSQ 23

RESULT 4
AAY65696
ID AAY65696 standard; Peptide; 34 AA.
XX
AC AAY65696;
XX
DT 10-FEB-2000 (first entry)
XX
DE TGF beta RII mutant peptide 1.
XX
KW Human; frameshift mutant; T cell response; tumour; treatment; cancer;
KW mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9958552-A2.
XX
PD 18-NOV-1999.
XX
PF 03-MAY-1999; 99WO-NO00143.
XX
PR 08-MAY-1998; 98NO-0002097.
XX
PA (NHYD) NORSK HYDRO AS.
XX
PI Gaudernack G, Eriksen JA, Moller M, Gjertsen MK, Saeterdal I;
XX
DR WPI; 2000-039064/03.
XX
PT New peptides derived from genes with frameshift mutations, used to
PT develop products for the treatment and prophylaxis of cancers -
XX
PS Claim 12; Page 20; 166pp; English.
XX
CC Peptides AAY65684-Y66142 are fragments of mutant proteins arising from a
CC frameshift mutation in a gene from a cancer cell. The peptides are
CC characterised in that they:
CC (i) are at least 8 amino acids long and a fragment of a mutant protein
CC arising from a frameshift mutation in a gene of a cancer cell;
CC (ii) consist of at least one amino acid of the mutant part of a protein
CC sequence encoded by the gene;
CC (iii) comprise 0-10 amino acid from the carboxyl terminus of the normal
CC part of the protein sequence preceding the amino terminus of the mutant
CC sequence and may further extend to the carboxyl terminus of the mutant
CC part of the protein as determined by a new stop codon generated by the
CC frameshift mutation; and
CC (iv) induce, either in their full lengths or after processing by an
CC antigen presenting cell (APC), T cell responses.
CC The genes that the peptides are derived from, are characterised as
CC susceptible to frameshift mutation by having a mono nucleoside base
CC repeat sequence of at least 5 residues, or a di-nucleoside base repeat
CC sequence of at least 4 di-nucleoside base units. The peptides are
CC created by the addition or deletion of 1 or 2 nucleoside base residues
CC from the repeat sequence. The novel peptides can elicit T cell responses
CC and toxicity against tumours and cancer cells carrying genes with
CC frameshift mutations. The novel peptides and DNA sequences can be used
CC for the preparation of a composition for the treatment or prophylaxis of
CC cancer.
XX
SQ Sequence 34 AA;

Query Match 100.0%; Score 106; DB 21; Length 34;
Best Local Similarity 100.0%; Pred. No. 4.3e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLVRLSSCVPVALMSAMTTSSSQ 23
||||||||||||||||||||||||
Db 1 SLVRLSSCVPVALMSAMTTSSSQ 23

RESULT 5
AAB82997
ID AAB82997 standard; Peptide; 34 AA.
XX
AC AAB82997;
XX
DT 21-DEC-2001 (first entry)
XX
DE Truncated TGF-beta receptor RII C-terminal sequence.
XX
KW Human; VACO457; transforming growth factor-beta receptor RII;
KW TGF-beta receptor RII; suppressor; tumour; colon cancer;
KW gastric cancer; breast cancer; diagnosis; gene therapy.
XX
OS Homo sapiens.
XX
PN US6291237-B1.
XX
PD 18-SEP-2001.
XX
PF 29-JAN-1999; 99US-0239864.
XX
PR 07-APR-1995; 95US-0417867.
PR 22-MAY-1995; 95US-0445520.
XX
PA (UYCA-) UNIV CASE WESTERN RESERVE.
PA (MEDI-) MEDICAL COLLEGE OHIO.
XX
PI Markowitz SD, Brattain MG, Willson JKV;
XX
DR WPI; 2001-637951/73.
XX
PT New isolated polynucleotides encoding a mutant form of transforming
PT growth factor beta receptor RII, useful in gene therapy, particularly
PT for treating cancers or tumours -
XX
PS Disclosure; Column 16; 30pp; English.
XX
CC The present sequence is that of the C-terminal region of a
CC truncated human transforming growth factor-beta receptor RII
CC (TGF-beta receptor RII) produced by colon cancer cell line VACO457.
CC In this cell line, the wild-type 10 bp polyadenine repeat (see
CC AAH27095) of the TGF-beta receptor RII gene is truncated by 1 base.
CC The mutant sequence encodes a truncated protein of 161 amino acids
CC (wild-type is 567 amino acids, see AAB82996), of which the last
CC 34 amino acids (present sequence) are altered from the wild-type,
CC which starting from Lys-128 has the sequence given in AAB82998.
CC Detection of RII mutant forms in tumour cell lines may be useful
CC for the development of a commercial test for RII mutation. The
CC invention is based on the discovery that the RII gene is a
CC cancer suppressor gene which is genetically inactivated (mutated)
CC in approximately 25% of colon cancers, including nearly all colon
CC cancers of the class identified as mutator/microsatellite
CC instability/RER. Methods for the diagnosis and prognosis of
CC cancer are based on detection of mutant forms of RII. Methods are

CC also provided for therapeutic intervention, including replacement
CC gene therapy.
XX
SQ Sequence 34 AA;

Query Match 100.0%; Score 106; DB 22; Length 34;
Best Local Similarity 100.0%; Pred. No. 4.3e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLVRLSSCVPVALMSAMTTSSSQ 23
||||||||||||||||||||||
Db 1 SLVRLSSCVPVALMSAMTTSSSQ 23

RESULT 6
ABB80865
ID ABB80865 standard; Protein; 34 AA.
XX
AC ABB80865;
XX
DT 08-OCT-2002 (first entry)
XX
DE Type II TGFbeta receptor (RII) mutant VACO457 C-terminal fragment.
XX
KW Transforming growth factor beta; TGFbeta; type II receptor; RII; RI;
KW tumour; cancer; cytostatic; gene therapy; immunotherapy; T cell therapy;
KW human; receptor; mutant.
XX
OS Homo sapiens.
XX
PN US2002064786-A1.
XX
PD 30-MAY-2002.
XX
PF 13-JUN-2001; 2001US-0878905.
XX
PR 29-JAN-1999; 99US-0239864.
PR 07-APR-1995; 95US-0417867.
PR 22-MAY-1995; 95US-0445520.
XX
PA (MARK/) MARKOWITZ S D.
PA (BRAT/) BRATTAIN M G.
PA (WILL/) WILLSON J K V.
XX
PI Markowitz SD, Brattain MG, Willson JKV;
XX
DR WPI; 2002-565743/60.
XX
PT Diagnosing cancer in patient comprises determining presence or absence
PT of functional type II receptor for transforming growth factor beta in
PT tissue from patient, the absence of functional RII receptor being
PT indicative of tumor tissue -
XX
PS Disclosure; Page 9; 30pp; English.
XX
CC The invention relates to diagnosing cancer in a patient by determining
CC presence or absence of functional type II receptor (RII) for transforming
CC growth factor beta (TGFbeta) in tissue from the patient, the absence of
CC functional RII being indicative of tumour tissue or precancerous lesions
CC in the patient. The methods are useful for diagnosing cancer in a
CC patient, predicting prognosis of a cancer patient, particularly a colon
CC cancer patient. Also in classifying tumour cell phenotype in a patient,
CC where the tumour tissue is chosen from colon cancer, endometrial cancer,
CC ovarian cancer, gastric cancer, pancreatic cancer and other malignancies,
CC and in treating colon cancer in a patient. The antibody specific to a
CC mutant protein of human TGF-beta receptor RII and an immunogenic
CC composition comprising the antibody, the non-functional mutant of the
CC growth regulatory gene product, or an expression vector encoding the same
CC non-functional mutant are useful for treating colon cancer in a patient,
CC where neoplastic cells of the patient express mutant form of RII. The
CC present sequence represents the C-terminal fragment of a RII receptor
CC mutant.
XX

SQ Sequence 34 AA;

Query Match 100.0%; Score 106; DB 23; Length 34;
Best Local Similarity 100.0%; Pred. No. 4.3e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLVRLSSCVPVALMSAMTTSSSQ 23
||||||||||||||||||||||
Db 1 SLVRLSSCVPVALMSAMTTSSSQ 23

RESULT 7
AAY65697
ID AAY65697 standard; Peptide; 44 AA.
XX
AC AAY65697;
XX
DT 10-FEB-2000 (first entry)
XX
DE TGF beta RII mutant peptide 2.
XX
KW Human; frameshift mutant; T cell response; tumour; treatment; cancer;
KW mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9958552-A2.
XX
PD 18-NOV-1999.
XX
PF 03-MAY-1999; 99WO-NO00143.
XX
PR 08-MAY-1998; 98NO-0002097.
XX
PA (NHYD) NORSK HYDRO AS.
XX
PI Gaudernack G, Eriksen JA, Moller M, Gjertsen MK, Saeterdal I;
XX
DR WPI; 2000-039064/03.
XX
PT New peptides derived from genes with frameshift mutations, used to
PT develop products for the treatment and prophylaxis of cancers -
XX
PS Claim 12; Page 20; 166pp; English.
XX
CC Peptides AAY65684-Y66142 are fragments of mutant proteins arising from a
CC frameshift mutation in a gene from a cancer cell. The peptides are
CC characterised in that they:
CC (i) are at least 8 amino acids long and a fragment of a mutant protein
CC arising from a frameshift mutation in a gene of a cancer cell;
CC (ii) consist of at least one amino acid of the mutant part of a protein
CC sequence encoded by the gene;
CC (iii) comprise 0-10 amino acid from the carboxyl terminus of the normal
CC part of the protein sequence preceding the amino terminus of the mutant
CC sequence and may further extend to the carboxyl terminus of the mutant
CC part of the protein as determined by a new stop codon generated by the
CC frameshift mutation; and
CC (iv) induce, either in their full lengths or after processing by an
CC antigen presenting cell (APC), T cell responses.
CC The genes that the peptides are derived from, are characterised as
CC susceptible to frameshift mutation by having a mono nucleoside base
CC repeat sequence of at least 5 residues, or a di-nucleoside base repeat
CC sequence of at least 4 di-nucleoside base units. The peptides are
CC created by the addition or deletion of 1 or 2 nucleoside base residues
CC from the repeat sequence. The novel peptides can elicit T cell responses
CC and toxicity against tumours and cancer cells carrying genes with
CC frameshift mutations. The novel peptides and DNA sequences can be used
CC for the preparation of a composition for the treatment or prophylaxis of
CC cancer.
XX
SQ Sequence 44 AA;

```
  Query Match             100.0%;  Score 106;  DB 21;  Length 44;
  Best Local Similarity   100.0%;  Pred. No. 5.8e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SLVRLSSCVPVALMSAMTTSSSQ 23
           |||||||||||||||||||||||
Db      11 SLVRLSSCVPVALMSAMTTSSSQ 33


RESULT 8
AAY66121
ID   AAY66121 standard; Peptide; 18 AA.
XX
AC   AAY66121;
XX
DT   10-FEB-2000  (first entry)
XX
DE   Frameshift mutated gene peptide 1.
XX
KW   Human; frameshift mutant; T cell response; tumour; treatment; cancer;
KW   mutein.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9958552-A2.
XX
PD   18-NOV-1999.
XX
PF   03-MAY-1999;   99WO-NO00143.
XX
PR   08-MAY-1998;   98NO-0002097.
XX
PA   (NHYD ) NORSK HYDRO AS.
XX
PI   Gaudernack G,  Eriksen JA,  Moller M,  Gjertsen MK,  Saeterdal I;
XX
DR   WPI; 2000-039064/03.
XX
PT   New peptides derived from genes with frameshift mutations, used to
PT   develop products for the treatment and prophylaxis of cancers -
XX
PS   Claim 12; Page 161; 166pp; English.
XX
CC   Peptides AAY65684-Y66142 are fragments of mutant proteins arising from a
CC   frameshift mutation in a gene from a cancer cell. The peptides are
CC   characterised in that they:
CC   (i) are at least 8 amino acids long and a fragment of a mutant protein
CC   arising from a frameshift mutation in a gene of a cancer cell;
CC   (ii) consist of at least one amino acid of the mutant part of a protein
CC   sequence encoded by the gene;
CC   (iii) comprise 0-10 amino acid from the carboxyl terminus of the normal
CC   part of the protein sequence preceding the amino terminus of the mutant
CC   sequence and may further extend to the carboxyl terminus of the mutant
CC   part of the protein as determined by a new stop codon generated by the
CC   frameshift mutation; and
CC   (iv) induce, either in their full lengths or after processing by an
CC   antigen presenting cell (APC), T cell responses.
CC   The genes that the peptides are derived from, are characterised as
CC   susceptible to frameshift mutation by having a mono nucleoside base
CC   repeat sequence of at least 5 residues, or a di-nucleoside base repeat
CC   sequence of at least 4 di-nucleoside base units. The peptides are
CC   created by the addition or deletion of 1 or 2 nucleoside base residues
CC   from the repeat sequence. The novel peptides can elicit T cell responses
CC   and toxicity against tumours and cancer cells carrying genes with
CC   frameshift mutations. The novel peptides and DNA sequences can be used
CC   for the preparation of a composition for the treatment or prophylaxis of
CC   cancer.
XX
SQ   Sequence   18 AA;

  Query Match             74.5%;  Score 79;  DB 21;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-06;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SLVRLSSCVPVALMSAM 17
           |||||||||||||||||
Db       2 SLVRLSSCVPVALMSAM 18


RESULT 9
AAY54038
ID   AAY54038 standard; Peptide; 26 AA.
XX
AC   AAY54038;
XX
DT   27-MAR-2000  (first entry)
XX
DE   Peptide used to produce antibodies.
XX
KW   Class I major histocompatibility glycoprotein complex; MHCI;
KW   mutant RII receptor; transforming growth factor-beta; TGF-beta;
KW   tumour; vaccine; gastric cancer; colon cancer; gene therapy.
XX
OS   Synthetic.
XX
PN   FR2779432-A1.
XX
PD   10-DEC-1999.
XX
PF   08-JUN-1998;   98FR-0007322.
XX
PR   08-JUN-1998;   98FR-0007322.
XX
PA   (TRGE ) TRANSGENE SA.
XX
DR   WPI; 2000-074958/07.
XX
PT   New nucleic acid sequences, useful for production of medicament for
PT   diagnosing, preventing and/or treating gastric or colon cancers -
XX
PS   Example 2; Page 30; 41pp; French.
XX
CC   The specification describes peptides which atatch themselves to at
CC   least one class I major histocompatibility glycoprotein complex (MHCI),
CC   and which do not comprise the sequence given in AAY54017. The peptides
CC   are derived from a mutant RII receptor of transforming growth factor-
CC   beta (TGF-beta). The presence of the mutant receptor leads to
CC   inactivation of TGF-beta, and contributes to the development of
CC   tumours. Especially, the mutation comprises the addition or deletion
CC   of an adenine between positions 709-718. The peptides, or nucleic
CC   acids encoding them, are useful for the production of a medicament
CC   (either preventative, therapeutic or as a vaccine) for treating gastric
CC   cancers or cancers of the colon by gene therapy or the peptide may be
CC   used as a diagnostic, prophylactic and/or therapeutic composition for
CC   the detection, prevention or treatment of gastric or colon cancers.
CC   The present sequence was used to raise antibodies for use in the course
CC   of the invention.
XX
SQ   Sequence   26 AA;

  Query Match             60.4%;  Score 64;  DB 21;  Length 26;
  Best Local Similarity   100.0%;  Pred. No. 0.002;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 PVALMSAMTTSSSQ 23
           ||||||||||||||
Db       4 PVALMSAMTTSSSQ 17


RESULT 10
AAY65701
ID   AAY65701 standard; Peptide; 22 AA.
XX
AC   AAY65701;
XX
```

DT 10-FEB-2000 (first entry)
XX
DE TGF beta RII mutant peptide 6.
XX
KW Human; frameshift mutant; T cell response; tumour; treatment; cancer;
KW mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9958552-A2.
XX
PD 18-NOV-1999.
XX
PF 03-MAY-1999; 99WO-NO00143.
XX
PR 08-MAY-1998; 98NO-0002097.
XX
PA (NHYD) NORSK HYDRO AS.
XX
PI Gaudernack G, Eriksen JA, Moller M, Gjertsen MK, Saeterdal I;
XX
DR WPI; 2000-039064/03.
XX
PT New peptides derived from genes with frameshift mutations, used to
PT develop products for the treatment and prophylaxis of cancers -
XX
PS Claim 12; Page 20; 166pp; English.
XX
CC Peptides AAY65684-Y66142 are fragments of mutant proteins arising from a
CC frameshift mutation in a gene from a cancer cell. The peptides are
CC characterised in that they:
CC (i) are at least 8 amino acids long and a fragment of a mutant protein
CC arising from a frameshift mutation in a gene of a cancer cell;
CC (ii) consist of at least one amino acid of the mutant part of a protein
CC sequence encoded by the gene;
CC (iii) comprise 0-10 amino acid from the carboxyl terminus of the normal
CC part of the protein sequence preceding the amino terminus of the mutant
CC sequence and may further extend to the carboxyl terminus of the mutant
CC part of the protein as determined by a new stop codon generated by the
CC frameshift mutation; and
CC (iv) induce, either in their full lengths or after processing by an
CC antigen presenting cell (APC), T cell responses.
CC The genes that the peptides are derived from, are characterised as
CC susceptible to frameshift mutation by having a mono nucleoside base
CC repeat sequence of at least 5 residues, or a di-nucleoside base repeat
CC sequence of at least 4 di-nucleoside base units. The peptides are
CC created by the addition or deletion of 1 or 2 nucleoside base residues
CC from the repeat sequence. The novel peptides can elicit T cell responses
CC and toxicity against tumours and cancer cells carrying genes with
CC frameshift mutations. The novel peptides and DNA sequences can be used
CC for the preparation of a composition for the treatment or prophylaxis of
CC cancer.
XX
SQ Sequence 22 AA;

Query Match 53.8%; Score 57; DB 21; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLVRLSSCVPVA 12
||||||||||||
Db 11 SLVRLSSCVPVA 22

RESULT 11
AAY54019
ID AAY54019 standard; peptide; 11 AA.
XX
AC AAY54019;
XX
DT 27-MAR-2000 (first entry)
XX
DE Peptide which is capable of binding MHCI glycoprotein HLA-A2.
XX
KW Class I major histocompatibility glycoprotein complex; MHCI;
KW mutant RII receptor; transforming growth factor-beta; TGF-beta;
KW tumour; vaccine; gastric cancer; colon cancer; gene therapy.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN FR2779432-A1.
XX
PD 10-DEC-1999.
XX
PF 08-JUN-1998; 98FR-0007322.
XX
PR 08-JUN-1998; 98FR-0007322.
XX
PA (TRGE) TRANSGENE SA.
XX
DR WPI; 2000-074958/07.
DR N-PSDB; AAZ37057.
XX
PT New nucleic acid sequences, useful for production of medicament for
PT diagnosing, preventing and/or treating gastric or colon cancers -
XX
PS Claim 2; Page 20; 41pp; French.
XX
CC The present sequence represents a peptide which is capable of fixing
CC itself on the glycoprotein HLA-A2 of the class I major
CC histocompatibility glycoprotein complex (MHCI). The specification
CC describes peptides which atatch themselves to at least one MHC1
CC glycoprotien, and which do not comprise the sequence given in AAY54017.
CC The peptides are derived from a mutant RII receptor of transforming
CC growth factor-beta (TGF-beta). The presence of the mutant receptor leads
CC to inactivation of TGF-beta, and contributes to the development of
CC tumours. Especially, the mutation comprises the addition or deletion of
CC an adenine between positions 709-718. The peptides, or nucleic acids
CC encoding them, are useful for the production of a medicament (either
CC preventative, therapeutic or as a vaccine) for treating gastric cancers
CC or cancers of the colon by gene therapy or the peptide may be used as a
CC diagnostic, prophylactic and/or therapeutic composition for the
CC detection, prevention or treatment of gastric or colon cancers.
XX
SQ Sequence 11 AA;

Query Match 50.0%; Score 53; DB 21; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.044;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLVRLSSCVPV 11
|||||||||||
Db 1 SLVRLSSCVPV 11

RESULT 12
AAY65699
ID AAY65699 standard; Peptide; 23 AA.
XX
AC AAY65699;
XX
DT 10-FEB-2000 (first entry)
XX
DE TGF beta RII mutant peptide 4.
XX
KW Human; frameshift mutant; T cell response; tumour; treatment; cancer;
KW mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9958552-A2.
XX
PD 18-NOV-1999.

```
XX
PF   03-MAY-1999;   99WO-NO00143.
XX
PR   08-MAY-1998;   98NO-0002097.
XX
PA   (NHYD ) NORSK HYDRO AS.
XX
PI   Gaudernack G,  Eriksen JA,  Moller M,  Gjertsen MK,  Saeterdal I;
XX
DR   WPI; 2000-039064/03.
XX
PT   New peptides derived from genes with frameshift mutations, used to
PT   develop products for the treatment and prophylaxis of cancers -
XX
PS   Claim 12; Page 20; 166pp; English.
XX
CC   Peptides AAY65684-Y66142 are fragments of mutant proteins arising from a
CC   frameshift mutation in a gene from a cancer cell. The peptides are
CC   characterised in that they:
CC   (i) are at least 8 amino acids long and a fragment of a mutant protein
CC   arising from a frameshift mutation in a gene of a cancer cell;
CC   (ii) consist of at least one amino acid of the mutant part of a protein
CC   sequence encoded by the gene;
CC   (iii) comprise 0-10 amino acid from the carboxyl terminus of the normal
CC   part of the protein sequence preceding the amino terminus of the mutant
CC   sequence and may further extend to the carboxyl terminus of the mutant
CC   part of the protein as determined by a new stop codon generated by the
CC   frameshift mutation; and
CC   (iv) induce, either in their full lengths or after processing by an
CC   antigen presenting cell (APC), T cell responses.
CC   The genes that the peptides are derived from, are characterised as
CC   susceptible to frameshift mutation by having a mono nucleoside base
CC   repeat sequence of at least 5 residues, or a di-nucleoside base repeat
CC   sequence of at least 4 di-nucleoside base units. The peptides are
CC   created by the addition or deletion of 1 or 2 nucleoside base residues
CC   from the repeat sequence. The novel peptides can elicit T cell responses
CC   and toxicity against tumours and cancer cells carrying genes with
CC   frameshift mutations. The novel peptides and DNA sequences can be used
CC   for the preparation of a composition for the treatment or prophylaxis of
CC   cancer.
XX
SQ   Sequence   23 AA;

  Query Match             50.0%;  Score 53;  DB 21;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 0.1;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     12 ALMSAMTTSSSQ 23
          ||||||||||||
Db      1 ALMSAMTTSSSQ 12


RESULT 13
AAY54022
ID   AAY54022 standard; peptide; 10 AA.
XX
AC   AAY54022;
XX
DT   27-MAR-2000  (first entry)
XX
DE   Peptide which is capable of binding MHCI glycoprotein HLA-A2.
XX
KW   Class I major histocompatibility glycoprotein complex; MHCI;
KW   mutant RII receptor; transforming growth factor-beta; TGF-beta;
KW   tumour; vaccine; gastric cancer; colon cancer; gene therapy.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   FR2779432-A1.
XX
PD   10-DEC-1999.
XX
PF   08-JUN-1998;   98FR-0007322.
XX
PR   08-JUN-1998;   98FR-0007322.
XX
PA   (TRGE ) TRANSGENE SA.
XX
DR   WPI; 2000-074958/07.
DR   N-PSDB; AAZ37060.
XX
PT   New nucleic acid sequences, useful for production of medicament for
PT   diagnosing, preventing and/or treating gastric or colon cancers -
XX
PS   Claim 2; Page 21; 41pp; French.
XX
CC   The present sequence represents a peptide which is capable of fixing
CC   itself on the glycoprotein HLA-A2 of the class I major
CC   histocompatibility glycoprotein complex (MHCI). The specification
CC   describes peptides which atatch themselves to at least one MHCl
CC   glycoprotien, and which do not comprise the sequence given in AAY54017.
CC   The peptides are derived from a mutant RII receptor of transforming
CC   growth factor-beta (TGF-beta). The presence of the mutant receptor leads
CC   to inactivation of TGF-beta, and contributes to the development of
CC   tumours. Especially, the mutation comprises the addition or deletion of
CC   an adenine between positions 709-718. The peptides, or nucleic acids
CC   encoding them, are useful for the production of a medicament (either
CC   preventative, therapeutic or as a vaccine) for treating gastric cancers
CC   or cancers of the colon by gene therapy or the peptide may be used as a
CC   diagnostic, prophylactic and/or therapeutic composition for the
CC   detection, prevention or treatment of gastric or colon cancers.
XX
SQ   Sequence   10 AA;

  Query Match             46.2%;  Score 49;  DB 21;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 0.18;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 RLSSCVPVAL 13
          ||||||||||
Db      1 RLSSCVPVAL 10


RESULT 14
AAG73473
ID   AAG73473 standard; Protein; 113 AA.
XX
AC   AAG73473;
XX
DT   10-AUG-2001  (first entry)
XX
DE   Human gene 17-encoded secreted protein fragment, SEQ ID NO:248.
XX
KW   Human; secreted protein; proliferative disorder; cancer; chromosome 2;
KW   foetal abnormality; developmental abnormality; haematopoietic disorder;
KW   immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW   inflammation; allergy; neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW   skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW   cardiovascular disorder; angiogenic disorder; kidney disorder;
KW   gastrointestinal disorder;  pregnancy-related disorder; tumour;
KW   endocrine disorder; infection; wound healing; vulnerary;
KW   cell culture; chemotaxis; food additive;
KW   binding partner identification.
XX
OS   Homo sapiens.
XX
PN   WO200134628-A1.
XX
PD   17-MAY-2001.
XX
PF   08-NOV-2000; 2000WO-US30653.
XX
PR   12-NOV-1999;   99US-0164735.
PR   27-JUL-2000; 2000US-0221193.
```

XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Komatsoulis GA, Birse CE, Ni J, Moore PA;
XX
DR WPI; 2001-329066/34.
XX
PT Nucleic acids encoding 35 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT disease and diabetic retinopathy -
XX
PS Disclosure; Page 39; 604pp; English.
XX
CC AAH32522-AAH32627 represent cDNAs corresponding to 35 human secreted
CC protein genes, and AAG73346-AAG73448 represent the proteins they encode.
CC AAG73449-AAG73519 represent human secreted protein fragments. The genes
CC and their corresponding secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 52 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinsons's disease), cognitive disorders, schizophrenia, asthma,
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein fragment referred to in the disclosure of the invention.
XX
SQ Sequence 113 AA;

Query Match 46.2%; Score 49; DB 22; Length 113;
Best Local Similarity 50.0%; Pred. No. 2.9;
Matches 10; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

```
Qy       3 VRLSSCVPVALMSAMTTSSS 22
           :| | | |||| :|:: |:|
Db      17 LRFSICSPVALRNALSRSTS 36
```

RESULT 15
AAU67640
ID AAU67640 standard; Protein; 138 AA.
XX
AC AAU67640;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #28536.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
DR N-PSDB; AAS59609.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
PS Example 1; SEQ ID No 28835; 1069pp; English.
XX
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 138 AA;

Query Match 43.4%; Score 46; DB 22; Length 138;
Best Local Similarity 45.5%; Pred. No. 11;
Matches 10; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

```
Qy       2 LVRLSSCVPVALMSAMTTSSSQ 23
           : ||| |||:|   |: |  |:
Db       2 VTRLSKCVPLACPPAIHTMCSK 23
```

Search completed: May 7, 2003, 09:30:03
Job time : 36 secs

OM protein - protein search, using sw model

Run on: May 7, 2003, 09:30:10 ; Search time 35 Seconds
(without alignments)
87.565 Million cell updates/sec

Title: US-09-674-973A-17
Perfect score: 23
Sequence: 1 SLVRLSSCVPVALMSAMTTSSSQ 23

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 8

Total number of hits satisfying chosen parameters: 35

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	23	100.0	23	21	AAY65700	TGF beta RII mutan
2	23	100.0	34	17	AAW05380	Fragment of VACO45
3	23	100.0	34	21	AAY54017	Peptide which is n
4	23	100.0	34	21	AAY65696	TGF beta RII mutan
5	23	100.0	34	22	AAB82997	Truncated TGF-beta
6	23	100.0	34	23	ABB80865	Type II TGFbeta re
7	23	100.0	44	21	AAY65697	TGF beta RII mutan
8	17	73.9	18	21	AAY66121	Frameshift mutated
9	14	60.9	26	21	AAY54038	Peptide used to pr
10	12	52.2	22	21	AAY65701	TGF beta RII mutan
11	12	52.2	23	21	AAY65699	TGF beta RII mutan
12	11	47.8	11	21	AAY54019	Peptide which is c
13	10	43.5	10	21	AAY54022	Peptide which is c
14	9	39.1	9	21	AAY54018	Peptide which is c
15	9	39.1	9	21	AAY54021	Peptide which is c
16	9	39.1	9	21	AAY54035	Peptide which is c
17	9	39.1	9	21	AAY54036	Peptide which is c
18	9	39.1	9	21	AAY54037	Peptide which is c
19	9	39.1	9	21	AAY66111	TGF beta RII mutan
20	9	39.1	9	21	AAY66122	Frameshift mutated
21	9	39.1	9	21	AAY66123	Frameshift mutated
22	9	39.1	9	21	AAY66124	Frameshift mutated
23	9	39.1	9	21	AAY66125	Frameshift mutated
24	9	39.1	9	21	AAY66126	Frameshift mutated
25	9	39.1	9	21	AAY66127	Frameshift mutated
26	9	39.1	9	21	AAY66128	Frameshift mutated
27	9	39.1	9	21	AAY66129	Frameshift mutated
28	9	39.1	10	21	AAY54025	Peptide capable of
29	9	39.1	10	21	AAY54029	Peptide which is c
30	9	39.1	19	21	AAY65698	TGF beta RII mutan
31	8	34.8	8	21	AAY54020	Peptide which is c
32	8	34.8	8	21	AAY54034	Peptide which is c
33	8	34.8	9	21	AAY54024	Peptide capable of
34	8	34.8	19	21	AAY65704	TGF beta RII mutan
35	8	34.8	35	20	AAY12053	Human 5' EST secre

ALIGNMENTS

RESULT 1
AAY65700
ID AAY65700 standard; Peptide; 23 AA.
XX
AC AAY65700;
XX
DT 10-FEB-2000 (first entry)
XX
DE TGF beta RII mutant peptide 5.
XX
KW Human; frameshift mutant; T cell response; tumour; treatment; cancer;
KW mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9958552-A2.
XX
PD 18-NOV-1999.
XX
PF 03-MAY-1998; 99WO-NO00143.
XX
PR 08-MAY-1998; 98NO-0002097.
XX
PA (NHYD) NORSK HYDRO AS.
XX
PI Gaudernack G, Eriksen JA, Moller M, Gjertsen MK, Saeterdal I;
XX
DR WPI; 2000-039064/03.
XX
PT New peptides derived from genes with frameshift mutations, used to
PT develop products for the treatment and prophylaxis of cancers -
XX
PS Claim 12; Page 20; 166pp; English.
XX
CC Peptides AAY65684-Y66142 are fragments of mutant proteins arising from a
CC frameshift mutation in a gene from a cancer cell. The peptides are
CC characterised in that they:
CC (i) are at least 8 amino acids long and a fragment of a mutant protein
CC arising from a frameshift mutation in a gene of a cancer cell;
CC (ii) consist of at least one amino acid of the mutant part of a protein
CC sequence encoded by the gene;
CC (iii) comprise 0-10 amino acid from the carboxyl terminus of the normal

CC part of the protein sequence preceding the amino terminus of the mutant
CC sequence and may further extend to the carboxyl terminus of the mutant
CC part of the protein as determined by a new stop codon generated by the
CC frameshift mutation; and
CC (iv) induce, either in their full lengths or after processing by an
CC antigen presenting cell (APC), T cell responses.
CC The genes that the peptides are derived from, are characterised as
CC susceptible to frameshift mutation by having a mono nucleoside base
CC repeat sequence of at least 5 residues, or a di-nucleoside base repeat
CC sequence of at least 4 di-nucleoside base units. The peptides are
CC created by the addition or deletion of 1 or 2 nucleoside base residues
CC from the repeat sequence. The novel peptides can elicit T cell responses
CC and toxicity against tumours and cancer cells carrying genes with
CC frameshift mutations. The novel peptides and DNA sequences can be used
CC for the preparation of a composition for the treatment or prophylaxis of
CC cancer.
XX
SQ Sequence 23 AA;

Query Match 100.0%; Score 23; DB 21; Length 23;
Best Local Similarity 100.0%; Pred. No. 8.2e-16;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLVRLSSCVPVALMSAMTTSSSQ 23
||||||||||||||||||||||||
Db 1 SLVRLSSCVPVALMSAMTTSSSQ 23

RESULT 2
AAW05380
ID AAW05380 standard; peptide; 34 AA.
XX
AC AAW05380;
XX
DT 04-JUN-1997 (first entry)
XX
DE Fragment of VACO457 RII mutant.
XX
KW Type I transforming growth factor beta receptor gene; epithelial cell;
KW tumour development; cancer; non-functional mutant; precancerous lesion;
KW growth regulatory gene; type II receptor; serine/threonine receptor;
KW tumour tissue; colonic cancer; endometrial cancer; ovarian cancer;
KW gastric cancer; TGFbeta receptor gene; pancreatic cancer.
XX
OS Synthetic.
XX
PN WO9631605-A1.
XX
PD 10-OCT-1996.
XX
PF 05-APR-1996; 96WO-US04727.
XX
PR 22-MAY-1995; 95US-0445520.
PR 07-APR-1995; 95US-0417867.
XX
PA (MEDI-) MEDICAL COLLEGE OHIO.
PA (UYCA-) UNIV CASE WESTERN.
XX
PI Brattain MG, Markowitz SD, Willson JKV;
XX
DR WPI; 1996-465028/46.
XX
PT Cancer diagnosis and therapy - based on mutation(s) in type II
PT transforming growth factor beta receptor
XX
PS Disclosure; Page 30; 70pp; English.
XX
CC This sequence represents a fragment of the type II transforming growth
CC factor beta (TGFbeta) receptor gene mutant VACO457. TGFbeta inhibits the
CC growth of multiple epithelial cell types, and loss of this negative
CC regulation is thought to contribute to tumour development. TGFbeta also
CC inhibits the growth of certain cancer cell lines. This sequence can be
CC detected by a method of the invention. The method of the invention is for
CC aiding cancer diagnosis or prognosis. The method comprises detecting
CC expression of a mutant form of type II TGFbeta receptor (mutant RII) by
CC cells of a patient or the absence of wild-type RII in tumour cells.
CC Another method comprises detecting a non-functional mutant form of a
CC growth regulatory gene which encodes a type II receptor which is a member
CC of a family of serine/threonine receptors that bind members of a family
CC of TGFbeta-like factors. Alternatively, the method comprises detecting a
CC mutant growth regulatory gene which contains repetitive DNA sequence
CC motifs in the wild-type coding region, where the presence of the
CC non-functional mutant form is indicative of tumour tissue or precancerous
CC lesions. The methods can be used for diagnosis or treatment of colonic,
CC endometrial, ovarian, gastric or pancreatic cancer or other malignancies.
XX
SQ Sequence 34 AA;

Query Match 100.0%; Score 23; DB 17; Length 34;
Best Local Similarity 100.0%; Pred. No. 1.2e-15;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLVRLSSCVPVALMSAMTTSSSQ 23
||||||||||||||||||||||||
Db 1 SLVRLSSCVPVALMSAMTTSSSQ 23

RESULT 3
AAY54017
ID AAY54017 standard; peptide; 34 AA.
XX
AC AAY54017;
XX
DT 27-MAR-2000 (first entry)
XX
DE Peptide which is not a part of MHCI glycoprotein binding peptides.
XX
KW Class I major histocompatibility glycoprotein complex; MHCI;
KW mutant RII receptor; transforming growth factor-beta; TGF-beta;
KW tumour; vaccine; gastric cancer; colon cancer; gene therapy.
XX
OS Homo sapiens.
XX
PN FR2779432-A1.
XX
PD 10-DEC-1999.
XX
PF 08-JUN-1998; 98FR-0007322.
XX
PR 08-JUN-1998; 98FR-0007322.
XX
PA (TRGE) TRANSGENE SA.
XX
DR WPI; 2000-074958/07.
XX
PT New nucleic acid sequences, useful for production of medicament for
PT diagnosing, preventing and/or treating gastric or colon cancers -
XX
PS Claim 1; Page 19; 41pp; French.
XX
CC The specification describes peptides which are capable of fixing
CC themselves on at least one class I major histocompatibility
CC glycoprotein complex (MHCI), and which do not comprise the present
CC sequence. The peptides are derived from a mutant RII receptor of
CC transforming growth factor-beta (TGF-beta). The presence of the
CC mutant receptor leads to inactivation of TGF-beta, and contributes
CC to the development of tumours. Especially, the mutation comprises
CC the addition or deletion of an adenine between positions 709-718.
CC The peptides, or nucleic acids encoding them, are useful for the
CC production of a medicament (either preventative, therapeutic or
CC as a vaccine) for treating gastric cancers or cancers of the colon
CC by gene therapy or the peptide may be used as a diagnostic,
CC prophylactic and/or therapeutic composition for the detection,
CC prevention or treatment of gastric or colon cancers.
XX
SQ Sequence 34 AA;

```
  Query Match             100.0%;  Score 23;  DB 21;  Length 34;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-15;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SLVRLSSCVPVALMSAMTTSSSQ 23
           |||||||||||||||||||||||
Db       1 SLVRLSSCVPVALMSAMTTSSSQ 23


RESULT 4
AAY65696
ID   AAY65696 standard; Peptide; 34 AA.
XX
AC   AAY65696;
XX
DT   10-FEB-2000  (first entry)
XX
DE   TGF beta RII mutant peptide 1.
XX
KW   Human; frameshift mutant; T cell response; tumour; treatment; cancer;
KW   mutein.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9958552-A2.
XX
PD   18-NOV-1999.
XX
PF   03-MAY-1999;   99WO-NO00143.
XX
PR   08-MAY-1998;   98NO-0002097.
XX
PA   (NHYD ) NORSK HYDRO AS.
XX
PI   Gaudernack G,  Eriksen JA,  Moller M,  Gjertsen MK,  Saeterdal I;
XX
DR   WPI; 2000-039064/03.
XX
PT   New peptides derived from genes with frameshift mutations, used to
PT   develop products for the treatment and prophylaxis of cancers -
XX
PS   Claim 12; Page 20; 166pp; English.
XX
CC   Peptides AAY65684-Y66142 are fragments of mutant proteins arising from a
CC   frameshift mutation in a gene from a cancer cell. The peptides are
CC   characterised in that they:
CC   (i) are at least 8 amino acids long and a fragment of a mutant protein
CC   arising from a frameshift mutation in a gene of a cancer cell;
CC   (ii) consist of at least one amino acid of the mutant part of a protein
CC   sequence encoded by the gene;
CC   (iii) comprise 0-10 amino acid from the carboxyl terminus of the normal
CC   part of the protein sequence preceding the amino terminus of the mutant
CC   sequence and may further extend to the carboxyl terminus of the mutant
CC   part of the protein as determined by a new stop codon generated by the
CC   frameshift mutation; and
CC   (iv) induce, either in their full lengths or after processing by an
CC   antigen presenting cell (APC), T cell responses.
CC   The genes that the peptides are derived from, are characterised as
CC   susceptible to frameshift mutation by having a mono nucleoside base
CC   repeat sequence of at least 5 residues, or a di-nucleoside base repeat
CC   sequence of at least 4 di-nucleoside base units. The peptides are
CC   created by the addition or deletion of 1 or 2 nucleoside base residues
CC   from the repeat sequence. The novel peptides can elicit T cell responses
CC   and toxicity against tumours and cancer cells carrying genes with
CC   frameshift mutations. The novel peptides and DNA sequences can be used
CC   for the preparation of a composition for the treatment or prophylaxis of
CC   cancer.
XX
SQ   Sequence   34 AA;

  Query Match             100.0%;  Score 23;  DB 21;  Length 34;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-15;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SLVRLSSCVPVALMSAMTTSSSQ 23
           |||||||||||||||||||||||
Db       1 SLVRLSSCVPVALMSAMTTSSSQ 23


RESULT 5
AAB82997
ID   AAB82997 standard; Peptide; 34 AA.
XX
AC   AAB82997;
XX
DT   21-DEC-2001  (first entry)
XX
DE   Truncated TGF-beta receptor RII C-terminal sequence.
XX
KW   Human; VACO457; transforming growth factor-beta receptor RII;
KW   TGF-beta receptor RII; suppressor; tumour; colon cancer;
KW   gastric cancer; breast cancer; diagnosis; gene therapy.
XX
OS   Homo sapiens.
XX
PN   US6291237-B1.
XX
PD   18-SEP-2001.
XX
PF   29-JAN-1999;   99US-0239864.
XX
PR   07-APR-1995;   95US-0417867.
PR   22-MAY-1995;   95US-0445520.
XX
PA   (UYCA-) UNIV CASE WESTERN RESERVE.
PA   (MEDI-) MEDICAL COLLEGE OHIO.
XX
PI   Markowitz SD,  Brattain MG,  Willson JKV;
XX
DR   WPI; 2001-637951/73.
XX
PT   New isolated polynucleotides encoding a mutant form of transforming
PT   growth factor beta receptor RII, useful in gene therapy, particularly
PT   for treating cancers or tumours -
XX
PS   Disclosure; Column 16; 30pp; English.
XX
CC   The present sequence is that of the C-terminal region of a
CC   truncated human transforming growth factor-beta receptor RII
CC   (TGF-beta receptor RII) produced by colon cancer cell line VACO457.
CC   In this cell line, the wild-type 10 bp polyadenine repeat (see
CC   AAH27095) of the TGF-beta receptor RII gene is truncated by 1 base.
CC   The mutant sequence encodes a truncated protein of 161 amino acids
CC   (wild-type is 567 amino acids, see AAB82996), of which the last
CC   34 amino acids (present sequence) are altered from the wild-type,
CC   which starting from Lys-128 has the sequence given in AAB82998.
CC   Detection of RII mutant forms in tumour cell lines may be useful
CC   for the development of a commercial test for RII mutation.  The
CC   invention is based on the discovery that the RII gene is a
CC   cancer suppressor gene which is genetically inactivated (mutated)
CC   in approximately 25% of colon cancers, including nearly all colon
CC   cancers of the class identified as mutator/microsatellite
CC   instability/RER.  Methods for the diagnosis and prognosis of
CC   cancer are based on detection of mutant forms of RII.  Methods are
CC   also provided for therapeutic intervention, including replacement
CC   gene therapy.
XX
SQ   Sequence   34 AA;

  Query Match             100.0%;  Score 23;  DB 22;  Length 34;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-15;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SLVRLSSCVPVALMSAMTTSSSQ 23
```

```
                    ||||||||||||||||||||||||
Db      1 SLVRLSSCVPVALMSAMTTSSSQ 23


RESULT 6
ABB80865
ID   ABB80865 standard; Protein; 34 AA.
XX
AC   ABB80865;
XX
DT   08-OCT-2002  (first entry)
XX
DE   Type II TGFbeta receptor (RII) mutant VACO457 C-terminal fragment.
XX
KW   Transforming growth factor beta; TGFbeta; type II receptor; RII; RI;
KW   tumour; cancer; cytostatic; gene therapy; immunotherapy; T cell therapy;
KW   human; receptor; mutant.
XX
OS   Homo sapiens.
XX
PN   US2002064786-A1.
XX
PD   30-MAY-2002.
XX
PF   13-JUN-2001; 2001US-0878905.
XX
PR   29-JAN-1999;   99US-0239864.
PR   07-APR-1995;   95US-0417867.
PR   22-MAY-1995;   95US-0445520.
XX
PA   (MARK/) MARKOWITZ S D.
PA   (BRAT/) BRATTAIN M G.
PA   (WILL/) WILLSON J K V.
XX
PI   Markowitz SD,  Brattain MG,  Willson JKV;
XX
DR   WPI; 2002-565743/60.
XX
PT   Diagnosing cancer in patient comprises determining presence or absence
PT   of functional type II receptor for transforming growth factor beta in
PT   tissue from patient, the absence of functional RII receptor being
PT   indicative of tumor tissue  -
XX
PS   Disclosure; Page 9; 30pp; English.
XX
CC   The invention relates to diagnosing cancer in a patient by determining
CC   presence or absence of functional type II receptor (RII) for transforming
CC   growth factor beta (TGFbeta) in tissue from the patient, the absence of
CC   functional RII being indicative of tumour tissue or precancerous lesions
CC   in the patient. The methods are useful for diagnosing cancer in a
CC   patient, predicting prognosis of a cancer patient, particularly a colon
CC   cancer patient. Also in classifying tumour cell phenotype in a patient,
CC   where the tumour tissue is chosen from colon cancer, endometrial cancer,
CC   ovarian cancer, gastric cancer, pancreatic cancer and other malignancies,
CC   and in treating colon cancer in a patient. The antibody specific to a
CC   mutant protein of human TGF-beta receptor RII and an immunogenic
CC   composition comprising the antibody, the non-functional mutant of the
CC   growth regulatory gene product, or an expression vector encoding the same
CC   non-functional mutant are useful for treating colon cancer in a patient,
CC   where neoplastic cells of the patient express mutant form of RII. The
CC   present sequence represents the C-terminal fragment of a RII receptor
CC   mutant.
XX
SQ   Sequence   34 AA;

  Query Match             100.0%;  Score 23;  DB 23;  Length 34;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-15;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SLVRLSSCVPVALMSAMTTSSSQ 23
          |||||||||||||||||||||||
Db      1 SLVRLSSCVPVALMSAMTTSSSQ 23


RESULT 7
AAY65697
ID   AAY65697 standard; Peptide; 44 AA.
XX
AC   AAY65697;
XX
DT   10-FEB-2000  (first entry)
XX
DE   TGF beta RII mutant peptide 2.
XX
KW   Human; frameshift mutant; T cell response; tumour; treatment; cancer;
KW   mutein.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9958552-A2.
XX
PD   18-NOV-1999.
XX
PF   03-MAY-1999;   99WO-NO00143.
XX
PR   08-MAY-1998;   98NO-0002097.
XX
PA   (NHYD ) NORSK HYDRO AS.
XX
PI   Gaudernack G,  Eriksen JA,  Moller M,  Gjertsen MK,  Saeterdal I;
XX
DR   WPI; 2000-039064/03.
XX
PT   New peptides derived from genes with frameshift mutations, used to
PT   develop products for the treatment and prophylaxis of cancers  -
XX
PS   Claim 12; Page 20; 166pp; English.
XX
CC   Peptides AAY65684-Y66142 are fragments of mutant proteins arising from a
CC   frameshift mutation in a gene from a cancer cell. The peptides are
CC   characterised in that they:
CC   (i) are at least 8 amino acids long and a fragment of a mutant protein
CC   arising from a frameshift mutation in a gene of a cancer cell;
CC   (ii) consist of at least one amino acid of the mutant part of a protein
CC   sequence encoded by the gene;
CC   (iii) comprise 0-10 amino acid from the carboxyl terminus of the normal
CC   part of the protein sequence preceding the amino terminus of the mutant
CC   sequence and may further extend to the carboxyl terminus of the mutant
CC   part of the protein as determined by a new stop codon generated by the
CC   frameshift mutation; and
CC   (iv) induce, either in their full lengths or after processing by an
CC   antigen presenting cell (APC), T cell responses.
CC   The genes that the peptides are derived from, are characterised as
CC   susceptible to frameshift mutation by having a mono nucleoside base
CC   repeat sequence of at least 5 residues, or a di-nucleoside base repeat
CC   sequence of at least 4 di-nucleoside base units. The peptides are
CC   created by the addition or deletion of 1 or 2 nucleoside base residues
CC   from the repeat sequence. The novel peptides can elicit T cell responses
CC   and toxicity against tumours and cancer cells carrying genes with
CC   frameshift mutations. The novel peptides and DNA sequences can be used
CC   for the preparation of a composition for the treatment or prophylaxis of
CC   cancer.
XX
SQ   Sequence   44 AA;

  Query Match             100.0%;  Score 23;  DB 21;  Length 44;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-15;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SLVRLSSCVPVALMSAMTTSSSQ 23
          |||||||||||||||||||||||
Db     11 SLVRLSSCVPVALMSAMTTSSSQ 33


RESULT 8
```

AAY66121
ID AAY66121 standard; Peptide; 18 AA.
XX
AC AAY66121;
XX
DT 10-FEB-2000 (first entry)
XX
DE Frameshift mutated gene peptide 1.
XX
KW Human; frameshift mutant; T cell response; tumour; treatment; cancer;
KW mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9958552-A2.
XX
PD 18-NOV-1999.
XX
PF 08-MAY-1999; 99WO-NO00143.
XX
PR 08-MAY-1998; 98NO-0002097.
XX
PA (NHYD) NORSK HYDRO AS.
XX
PI Gaudernack G, Eriksen JA, Moller M, Gjertsen MK, Saeterdal I;
XX
DR WPI; 2000-039064/03.
XX
PT New peptides derived from genes with frameshift mutations, used to
PT develop products for the treatment and prophylaxis of cancers -
XX
PS Claim 12; Page 161; 166pp; English.
XX
CC Peptides AAY65684-Y66142 are fragments of mutant proteins arising from a
CC frameshift mutation in a gene from a cancer cell. The peptides are
CC characterised in that they:
CC (i) are at least 8 amino acids long and a fragment of a mutant protein
CC arising from a frameshift mutation in a gene of a cancer cell;
CC (ii) consist of at least one amino acid of the mutant part of a protein
CC sequence encoded by the gene;
CC (iii) comprise 0-10 amino acid from the carboxyl terminus of the normal
CC part of the protein sequence preceding the amino terminus of the mutant
CC sequence and may further extend to the carboxyl terminus of the mutant
CC part of the protein as determined by a new stop codon generated by the
CC frameshift mutation; and
CC (iv) induce, either in their full lengths or after processing by an
CC antigen presenting cell (APC), T cell responses.
CC The genes that the peptides are derived from, are characterised as
CC susceptible to frameshift mutation by having a mono nucleoside base
CC repeat sequence of at least 5 residues, or a di-nucleoside base repeat
CC sequence of at least 4 di-nucleoside base units. The peptides are
CC created by the addition or deletion of 1 or 2 nucleoside base residues
CC from the repeat sequence. The novel peptides can elicit T cell responses
CC and toxicity against tumours and cancer cells carrying genes with
CC frameshift mutations. The novel peptides and DNA sequences can be used
CC for the preparation of a composition for the treatment or prophylaxis of
CC cancer.
XX
SQ Sequence 18 AA;

Query Match 73.9%; Score 17; DB 21; Length 18;
Best Local Similarity 100.0%; Pred. No. 4.8e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLVRLSSCVPVALMSAM 17
||||||||||||||||| |
Db 2 SLVRLSSCVPVALMSAM 18

RESULT 9
AAY54038
ID AAY54038 standard; Peptide; 26 AA.
XX
AC AAY54038;
XX
DT 27-MAR-2000 (first entry)
XX
DE Peptide used to produce antibodies.
XX
KW Class I major histocompatibility glycoprotein complex; MHCI;
KW mutant RII receptor; transforming growth factor-beta; TGF-beta;
KW tumour; vaccine; gastric cancer; colon cancer; gene therapy.
XX
OS Synthetic.
XX
PN FR2779432-A1.
XX
PD 10-DEC-1999.
XX
PF 08-JUN-1998; 98FR-0007322.
XX
PR 08-JUN-1998; 98FR-0007322.
XX
PA (TRGE) TRANSGENE SA.
XX
DR WPI; 2000-074958/07.
XX
PT New nucleic acid sequences, useful for production of medicament for
PT diagnosing, preventing and/or treating gastric or colon cancers -
XX
PS Example 2; Page 30; 41pp; French.
XX
CC The specification describes peptides which atatch themselves to at
CC least one class I major histocompatibility glycoprotein complex (MHCI),
CC and which do not comprise the sequence given in AAY54017. The peptides
CC are derived from a mutant RII receptor of transforming growth factor-
CC beta (TGF-beta). The presence of the mutant receptor leads to
CC inactivation of TGF-beta, and contributes to the development of
CC tumours. Especially, the mutation comprises the addition or deletion
CC of an adenine between positions 709-718. The peptides, or nucleic
CC acids encoding them, are useful for the production of a medicament
CC (either preventative, therapeutic or as a vaccine) for treating gastric
CC cancers or cancers of the colon by gene therapy or the peptide may be
CC used as a diagnostic, prophylactic and/or therapeutic composition for
CC the detection, prevention or treatment of gastric or colon cancers.
CC The present sequence was used to raise antibodies for use in the course
CC of the invention.
XX
SQ Sequence 26 AA;

Query Match 60.9%; Score 14; DB 21; Length 26;
Best Local Similarity 100.0%; Pred. No. 5.8e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 PVALMSAMTTSSSQ 23
|||||||||||||| |
Db 4 PVALMSAMTTSSSQ 17

RESULT 10
AAY65701
ID AAY65701 standard; Peptide; 22 AA.
XX
AC AAY65701;
XX
DT 10-FEB-2000 (first entry)
XX
DE TGF beta RII mutant peptide 6.
XX
KW Human; frameshift mutant; T cell response; tumour; treatment; cancer;
KW mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX

PN WO9958552-A2.
XX
PD 18-NOV-1999.
XX
PF 03-MAY-1999; 99WO-NO00143.
XX
PR 08-MAY-1998; 98NO-0002097.
XX
PA (NHYD) NORSK HYDRO AS.
XX
PI Gaudernack G, Eriksen JA, Moller M, Gjertsen MK, Saeterdal I;
XX
DR WPI; 2000-039064/03.
XX
PT New peptides derived from genes with frameshift mutations, used to
PT develop products for the treatment and prophylaxis of cancers -
XX
PS Claim 12; Page 20; 166pp; English.
XX
CC Peptides AAY65684-Y66142 are fragments of mutant proteins arising from a
CC frameshift mutation in a gene from a cancer cell. The peptides are
CC characterised in that they:
CC (i) are at least 8 amino acids long and a fragment of a mutant protein
CC arising from a frameshift mutation in a gene of a cancer cell;
CC (ii) consist of at least one amino acid of the mutant part of a protein
CC sequence encoded by the gene;
CC (iii) comprise 0-10 amino acid from the carboxyl terminus of the normal
CC part of the protein sequence preceding the amino terminus of the mutant
CC sequence and may further extend to the carboxyl terminus of the mutant
CC part of the protein as determined by a new stop codon generated by the
CC frameshift mutation; and
CC (iv) induce, either in their full lengths or after processing by an
CC antigen presenting cell (APC), T cell responses.
CC The genes that the peptides are derived from, are characterised as
CC susceptible to frameshift mutation by having a mono nucleoside base
CC repeat sequence of at least 5 residues, or a di-nucleoside base repeat
CC sequence of at least 4 di-nucleoside base units. The peptides are
CC created by the addition or deletion of 1 or 2 nucleoside base residues
CC from the repeat sequence. The novel peptides can elicit T cell responses
CC and toxicity against tumours and cancer cells carrying genes with
CC frameshift mutations. The novel peptides and DNA sequences can be used
CC for the preparation of a composition for the treatment or prophylaxis of
CC cancer.
XX
SQ Sequence 22 AA;

Query Match 52.2%; Score 12; DB 21; Length 22;
Best Local Similarity 100.0%; Pred. No. 4.5e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLVRLSSCVPVA 12
||||||||||||
Db 11 SLVRLSSCVPVA 22

RESULT 11
AAY65699
ID AAY65699 standard; Peptide; 23 AA.
XX
AC AAY65699;
XX
DT 10-FEB-2000 (first entry)
XX
DE TGF beta RII mutant peptide 4.
XX
KW Human; frameshift mutant; T cell response; tumour; treatment; cancer;
KW mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9958552-A2.
XX
PD 18-NOV-1999.
XX
PF 03-MAY-1999; 99WO-NO00143.
XX
PR 08-MAY-1998; 98NO-0002097.
XX
PA (NHYD) NORSK HYDRO AS.
XX
PI Gaudernack G, Eriksen JA, Moller M, Gjertsen MK, Saeterdal I;
XX
DR WPI; 2000-039064/03.
XX
PT New peptides derived from genes with frameshift mutations, used to
PT develop products for the treatment and prophylaxis of cancers -
XX
PS Claim 12; Page 20; 166pp; English.
XX
CC Peptides AAY65684-Y66142 are fragments of mutant proteins arising from a
CC frameshift mutation in a gene from a cancer cell. The peptides are
CC characterised in that they:
CC (i) are at least 8 amino acids long and a fragment of a mutant protein
CC arising from a frameshift mutation in a gene of a cancer cell;
CC (ii) consist of at least one amino acid of the mutant part of a protein
CC sequence encoded by the gene;
CC (iii) comprise 0-10 amino acid from the carboxyl terminus of the normal
CC part of the protein sequence preceding the amino terminus of the mutant
CC sequence and may further extend to the carboxyl terminus of the mutant
CC part of the protein as determined by a new stop codon generated by the
CC frameshift mutation; and
CC (iv) induce, either in their full lengths or after processing by an
CC antigen presenting cell (APC), T cell responses.
CC The genes that the peptides are derived from, are characterised as
CC susceptible to frameshift mutation by having a mono nucleoside base
CC repeat sequence of at least 5 residues, or a di-nucleoside base repeat
CC sequence of at least 4 di-nucleoside base units. The peptides are
CC created by the addition or deletion of 1 or 2 nucleoside base residues
CC from the repeat sequence. The novel peptides can elicit T cell responses
CC and toxicity against tumours and cancer cells carrying genes with
CC frameshift mutations. The novel peptides and DNA sequences can be used
CC for the preparation of a composition for the treatment or prophylaxis of
CC cancer.
XX
SQ Sequence 23 AA;

Query Match 52.2%; Score 12; DB 21; Length 23;
Best Local Similarity 100.0%; Pred. No. 4.6e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 ALMSAMTTSSSQ 23
||||||||||||
Db 1 ALMSAMTTSSSQ 12

RESULT 12
AAY54019
ID AAY54019 standard; peptide; 11 AA.
XX
AC AAY54019;
XX
DT 27-MAR-2000 (first entry)
XX
DE Peptide which is capable of binding MHCI glycoprotein HLA-A2.
XX
KW Class I major histocompatibility glycoprotein complex; MHCI;
KW mutant RII receptor; transforming growth factor-beta; TGF-beta;
KW tumour; vaccine; gastric cancer; colon cancer; gene therapy.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN FR2779432-A1.
XX
PD 10-DEC-1999.

XX
PF 08-JUN-1998; 98FR-0007322.
XX
PR 08-JUN-1998; 98FR-0007322.
XX
PA (TRGE) TRANSGENE SA.
XX
DR WPI; 2000-074958/07.
DR N-PSDB; AAZ37057.
XX
PT New nucleic acid sequences, useful for production of medicament for
PT diagnosing, preventing and/or treating gastric or colon cancers -
XX
PS Claim 2; Page 20; 41pp; French.
XX
CC The present sequence represents a peptide which is capable of fixing
CC itself on the glycoprotein HLA-A2 of the class I major
CC histocompatibility glycoprotein complex (MHCI). The specification
CC describes peptides which atatch themselves to at least one MHC1
CC glycoprotien, and which do not comprise the sequence given in AAY54017.
CC The peptides are derived from a mutant RII receptor of transforming
CC growth factor-beta (TGF-beta). The presence of the mutant receptor leads
CC to inactivation of TGF-beta, and contributes to the development of
CC tumours. Especially, the mutation comprises the addition or deletion of
CC an adenine between positions 709-718. The peptides, or nucleic acids
CC encoding them, are useful for the production of a medicament (either
CC preventative, therapeutic or as a vaccine) for treating gastric cancers
CC or cancers of the colon by gene therapy or the peptide may be used as a
CC diagnostic, prophylactic and/or therapeutic composition for the
CC detection, prevention or treatment of gastric or colon cancers.
XX
SQ Sequence 11 AA;

Query Match 47.8%; Score 11; DB 21; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.00022;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLVRLSSCVPV 11
||||||||||||
Db 1 SLVRLSSCVPV 11

RESULT 13
AAY54022
ID AAY54022 standard; peptide; 10 AA.
XX
AC AAY54022;
XX
DT 27-MAR-2000 (first entry)
XX
DE Peptide which is capable of binding MHCI glycoprotein HLA-A2.
XX
KW Class I major histocompatibility glycoprotein complex; MHCI;
KW mutant RII receptor; transforming growth factor-beta; TGF-beta;
KW tumour; vaccine; gastric cancer; colon cancer; gene therapy.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN FR2779432-A1.
XX
PD 10-DEC-1999.
XX
PF 08-JUN-1998; 98FR-0007322.
XX
PR 08-JUN-1998; 98FR-0007322.
XX
PA (TRGE) TRANSGENE SA.
XX
DR WPI; 2000-074958/07.
DR N-PSDB; AAZ37060.
XX
PT New nucleic acid sequences, useful for production of medicament for
PT diagnosing, preventing and/or treating gastric or colon cancers -
XX
PS Claim 2; Page 21; 41pp; French.
XX
CC The present sequence represents a peptide which is capable of fixing
CC itself on the glycoprotein HLA-A2 of the class I major
CC histocompatibility glycoprotein complex (MHCI). The specification
CC describes peptides which atatch themselves to at least one MHC1
CC glycoprotien, and which do not comprise the sequence given in AAY54017.
CC The peptides are derived from a mutant RII receptor of transforming
CC growth factor-beta (TGF-beta). The presence of the mutant receptor leads
CC to inactivation of TGF-beta, and contributes to the development of
CC tumours. Especially, the mutation comprises the addition or deletion of
CC an adenine between positions 709-718. The peptides, or nucleic acids
CC encoding them, are useful for the production of a medicament (either
CC preventative, therapeutic or as a vaccine) for treating gastric cancers
CC or cancers of the colon by gene therapy or the peptide may be used as a
CC diagnostic, prophylactic and/or therapeutic composition for the
CC detection, prevention or treatment of gastric or colon cancers.
XX
SQ Sequence 10 AA;

Query Match 43.5%; Score 10; DB 21; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 RLSSCVPVAL 13
||||||||||
Db 1 RLSSCVPVAL 10

RESULT 14
AAY54018
ID AAY54018 standard; peptide; 9 AA.
XX
AC AAY54018;
XX
DT 27-MAR-2000 (first entry)
XX
DE Peptide which is capable of binding MHCI glycoprotein HLA-A2.
XX
KW Class I major histocompatibility glycoprotein complex; MHCI;
KW mutant RII receptor; transforming growth factor-beta; TGF-beta;
KW tumour; vaccine; gastric cancer; colon cancer; gene therapy.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN FR2779432-A1.
XX
PD 10-DEC-1999.
XX
PF 08-JUN-1998; 98FR-0007322.
XX
PR 08-JUN-1998; 98FR-0007322.
XX
PA (TRGE) TRANSGENE SA.
XX
DR WPI; 2000-074958/07.
DR N-PSDB; AAZ37056.
XX
PT New nucleic acid sequences, useful for production of medicament for
PT diagnosing, preventing and/or treating gastric or colon cancers -
XX
PS Claim 2; Page 20; 41pp; French.
XX
CC The present sequence represents a peptide which is capable of fixing
CC itself on the glycoprotein HLA-A2 of the class I major
CC histocompatibility glycoprotein complex (MHCI). The specification
CC describes peptides which atatch themselves to at least one MHC1
CC glycoprotien, and which do not comprise the sequence given in AAY54017.
CC The peptides are derived from a mutant RII receptor of transforming
CC growth factor-beta (TGF-beta). The presence of the mutant receptor leads

CC to inactivation of TGF-beta, and contributes to the development of
CC tumours. Especially, the mutation comprises the addition or deletion of
CC an adenine between positions 709-718. The peptides, or nucleic acids
CC encoding them, are useful for the production of a medicament (either
CC preventative, therapeutic or as a vaccine) for treating gastric cancers
CC or cancers of the colon by gene therapy or the peptide may be used as a
CC diagnostic, prophylactic and/or therapeutic composition for the
CC detection, prevention or treatment of gastric or colon cancers.
XX
SQ Sequence 9 AA;

Query Match 39.1%; Score 9; DB 21; Length 9;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLVRLSSCV 9
|||||||||
Db 1 SLVRLSSCV 9

RESULT 15
AAY54021
ID AAY54021 standard; peptide; 9 AA.
XX
AC AAY54021;
XX
DT 27-MAR-2000 (first entry)
XX
DE Peptide which is capable of binding MHCI glycoprotein HLA-A2.
XX
KW Class I major histocompatibility glycoprotein complex; MHCI;
KW mutant RII receptor; transforming growth factor-beta; TGF-beta;
KW tumour; vaccine; gastric cancer; colon cancer; gene therapy.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN FR2779432-A1.
XX
PD 10-DEC-1999.
XX
PF 08-JUN-1998; 98FR-0007322.
XX
PR 08-JUN-1998; 98FR-0007322.
XX
PA (TRGE) TRANSGENE SA.
XX
DR WPI; 2000-074958/07.
DR N-PSDB; AAZ37059.
XX
PT New nucleic acid sequences, useful for production of medicament for
PT diagnosing, preventing and/or treating gastric or colon cancers -
XX
PS Claim 2; Page 20; 41pp; French.
XX
CC The present sequence represents a peptide which is capable of fixing
CC itself on the glycoprotein HLA-A2 of the class I major
CC histocompatibility glycoprotein complex (MHCI). The specification
CC describes peptides which atatch themselves to at least one MHCl
CC glycoprotien, and which do not comprise the sequence given in AAY54017.
CC The peptides are derived from a mutant RII receptor of transforming
CC growth factor-beta (TGF-beta). The presence of the mutant receptor leads
CC to inactivation of TGF-beta, and contributes to the development of
CC tumours. Especially, the mutation comprises the addition or deletion of
CC an adenine between positions 709-718. The peptides, or nucleic acids
CC encoding them, are useful for the production of a medicament (either
CC preventative, therapeutic or as a vaccine) for treating gastric cancers
CC or cancers of the colon by gene therapy or the peptide may be used as a
CC diagnostic, prophylactic and/or therapeutic composition for the
CC detection, prevention or treatment of gastric or colon cancers.
XX
SQ Sequence 9 AA;

Query Match 39.1%; Score 9; DB 21; Length 9;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 RLSSCVPVA 12
|||||||||
Db 1 RLSSCVPVA 9

RESULT 16
AAY54035
ID AAY54035 standard; peptide; 9 AA.
XX
AC AAY54035;
XX
DT 27-MAR-2000 (first entry)
XX
DE Peptide which is capable of binding MHCI glycoprotein HLA-B7.
XX
KW Class I major histocompatibility glycoprotein complex; MHCI;
KW mutant RII receptor; transforming growth factor-beta; TGF-beta;
KW tumour; vaccine; gastric cancer; colon cancer; gene therapy.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN FR2779432-A1.
XX
PD 10-DEC-1999.
XX
PF 08-JUN-1998; 98FR-0007322.
XX
PR 08-JUN-1998; 98FR-0007322.
XX
PA (TRGE) TRANSGENE SA.
XX
DR WPI; 2000-074958/07.
DR N-PSDB; AAZ37073.
XX
PT New nucleic acid sequences, useful for production of medicament for
PT diagnosing, preventing and/or treating gastric or colon cancers -
XX
PS Claim 2; Page 24; 41pp; French.
XX
CC The present sequence represents a peptide which is capable of fixing
CC itself on the glycoprotein HLA-B7 of the class I major
CC histocompatibility glycoprotein complex (MHCI). The specification
CC describes peptides which atatch themselves to at least one MHCl
CC glycoprotien, and which do not comprise the sequence given in AAY54017.
CC The peptides are derived from a mutant RII receptor of transforming
CC growth factor-beta (TGF-beta). The presence of the mutant receptor leads
CC to inactivation of TGF-beta, and contributes to the development of
CC tumours. Especially, the mutation comprises the addition or deletion of
CC an adenine between positions 709-718. The peptides, or nucleic acids
CC encoding them, are useful for the production of a medicament (either
CC preventative, therapeutic or as a vaccine) for treating gastric cancers
CC or cancers of the colon by gene therapy or the peptide may be used as a
CC diagnostic, prophylactic and/or therapeutic composition for the
CC detection, prevention or treatment of gastric or colon cancers.
XX
SQ Sequence 9 AA;

Query Match 39.1%; Score 9; DB 21; Length 9;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LSSCVPVAL 13
|||||||||
Db 1 LSSCVPVAL 9

RESULT 17
AAY54036

ID AAY54036 standard; peptide; 9 AA.
XX
AC AAY54036;
XX
DT 27-MAR-2000 (first entry)
XX
DE Peptide which is capable of binding MHCI glycoprotein HLA-B35.
XX
KW Class I major histocompatibility glycoprotein complex; MHCI;
KW mutant RII receptor; transforming growth factor-beta; TGF-beta;
KW tumour; vaccine; gastric cancer; colon cancer; gene therapy.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN FR2779432-A1.
XX
PD 10-DEC-1999.
XX
PF 08-JUN-1998; 98FR-0007322.
XX
PR 08-JUN-1998; 98FR-0007322.
XX
PA (TRGE) TRANSGENE SA.
XX
DR WPI; 2000-074958/07.
DR N-PSDB; AAZ37074.
XX
PT New nucleic acid sequences, useful for production of medicament for
PT diagnosing, preventing and/or treating gastric or colon cancers -
XX
PS Claim 2; Page 24; 41pp; French.
XX
CC The present sequence represents a peptide which is capable of fixing
CC itself on glycoprotein HLA-B35 of the class I major histocompatibility
CC glycoprotein complex (MHCI). The specification describes peptides
CC which atatch themselves to at least one MHC1 glycoprotien, and which do
CC not comprise the sequence given in AAY54017. The peptides are derived
CC from a mutant RII receptor of transforming growth factor-beta
CC (TGF-beta). The presence of the mutant receptor leads to inactivation of
CC TGF-beta, and contributes to the development of tumours. Especially, the
CC mutation comprises the addition or deletion of an adenine between
CC positions 709-718. The peptides, or nucleic acids encoding them, are
CC useful for the production of a medicament (either preventative,
CC therapeutic or as a vaccine) for treating gastric cancers or cancers of
CC the colon by gene therapy or the peptide may be used as a diagnostic,
CC prophylactic and/or therapeutic composition for the detection,
CC prevention or treatment of gastric or colon cancers.
XX
SQ Sequence 9 AA;

Query Match 39.1%; Score 9; DB 21; Length 9;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 VPVALMSAM 17
||||||||| |
Db 1 VPVALMSAM 9

RESULT 18
AAY54037
ID AAY54037 standard; peptide; 9 AA.
XX
AC AAY54037;
XX
DT 27-MAR-2000 (first entry)
XX
DE Peptide which is capable of binding MHCI glycoprotein HLA-B27.
XX
KW Class I major histocompatibility glycoprotein complex; MHCI;
KW mutant RII receptor; transforming growth factor-beta; TGF-beta;
KW tumour; vaccine; gastric cancer; colon cancer; gene therapy.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN FR2779432-A1.
XX
PD 10-DEC-1999.
XX
PF 08-JUN-1998; 98FR-0007322.
XX
PR 08-JUN-1998; 98FR-0007322.
XX
PA (TRGE) TRANSGENE SA.
XX
DR WPI; 2000-074958/07.
DR N-PSDB; AAZ37075.
XX
PT New nucleic acid sequences, useful for production of medicament for
PT diagnosing, preventing and/or treating gastric or colon cancers -
XX
PS Claim 2; Page 24; 41pp; French.
XX
CC The present sequence represents a peptide which is capable of fixing
CC itself on glycoprotein HLA-B27 of the class I major histocompatibility
CC glycoprotein complex (MHCI). The specification describes peptides
CC which atatch themselves to at least one MHC1 glycoprotien, and which do
CC not comprise the sequence given in AAY54017. The peptides are derived
CC from a mutant RII receptor of transforming growth factor-beta
CC (TGF-beta). The presence of the mutant receptor leads to inactivation of
CC TGF-beta, and contributes to the development of tumours. Especially, the
CC mutation comprises the addition or deletion of an adenine between
CC positions 709-718. The peptides, or nucleic acids encoding them, are
CC useful for the production of a medicament (either preventative,
CC therapeutic or as a vaccine) for treating gastric cancers or cancers of
CC the colon by gene therapy or the peptide may be used as a diagnostic,
CC prophylactic and/or therapeutic composition for the detection,
CC prevention or treatment of gastric or colon cancers.
XX
SQ Sequence 9 AA;

Query Match 39.1%; Score 9; DB 21; Length 9;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 VRLSSCVPV 11
||||||||| |
Db 1 VRLSSCVPV 9

RESULT 19
AAY66111
ID AAY66111 standard; Peptide; 9 AA.
XX
AC AAY66111;
XX
DT 10-FEB-2000 (first entry)
XX
DE TGF beta RII mutant peptide 10.
XX
KW Human; frameshift mutant; T cell response; tumour; treatment; cancer;
KW mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9958552-A2.
XX
PD 18-NOV-1999.
XX
PF 03-MAY-1999; 99WO-NO00143.
XX
PR 08-MAY-1998; 98NO-0002097.
XX

PA (NHYD) NORSK HYDRO AS.
XX
PI Gaudernack G, Eriksen JA, Moller M, Gjertsen MK, Saeterdal I;
XX
DR WPI; 2000-039064/03.
XX
PT New peptides derived from genes with frameshift mutations, used to
PT develop products for the treatment and prophylaxis of cancers -
XX
PS Claim 12; Page 20; 166pp; English.
XX
CC Peptides AAY65684-Y66142 are fragments of mutant proteins arising from a
CC frameshift mutation in a gene from a cancer cell. The peptides are
CC characterised in that they:
CC (i) are at least 8 amino acids long and a fragment of a mutant protein
CC arising from a frameshift mutation in a gene of a cancer cell;
CC (ii) consist of at least one amino acid of the mutant part of a protein
CC sequence encoded by the gene;
CC (iii) comprise 0-10 amino acid from the carboxyl terminus of the normal
CC part of the protein sequence preceding the amino terminus of the mutant
CC sequence and may further extend to the carboxyl terminus of the mutant
CC part of the protein as determined by a new stop codon generated by the
CC frameshift mutation; and
CC (iv) induce, either in their full lengths or after processing by an
CC antigen presenting cell (APC), T cell responses.
CC The genes that the peptides are derived from, are characterised as
CC susceptible to frameshift mutation by having a mono nucleoside base
CC repeat sequence of at least 5 residues, or a di-nucleoside base repeat
CC sequence of at least 4 di-nucleoside base units. The peptides are
CC created by the addition or deletion of 1 or 2 nucleoside base residues
CC from the repeat sequence. The novel peptides can elicit T cell responses
CC and toxicity against tumours and cancer cells carrying genes with
CC frameshift mutations. The novel peptides and DNA sequences can be used
CC for the preparation of a composition for the treatment or prophylaxis of
CC cancer.
XX
SQ Sequence 9 AA;

Query Match 39.1%; Score 9; DB 21; Length 9;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLVRLSSCV 9
|||||||||
Db 1 SLVRLSSCV 9

RESULT 20
AAY66122
ID AAY66122 standard; Peptide; 9 AA.
XX
AC AAY66122;
XX
DT 10-FEB-2000 (first entry)
XX
DE Frameshift mutated gene peptide 2.
XX
KW Human; frameshift mutant; T cell response; tumour; treatment; cancer;
KW mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9958552-A2.
XX
PD 18-NOV-1999.
XX
PF 03-MAY-1999; 99WO-NO00143.
XX
PR 08-MAY-1998; 98NO-0002097.
XX
PA (NHYD) NORSK HYDRO AS.
XX
PI Gaudernack G, Eriksen JA, Moller M, Gjertsen MK, Saeterdal I;
XX
DR WPI; 2000-039064/03.
XX
PT New peptides derived from genes with frameshift mutations, used to
PT develop products for the treatment and prophylaxis of cancers -
XX
PS Claim 13; Page 161; 166pp; English.
XX
CC Peptides AAY65684-Y66142 are fragments of mutant proteins arising from a
CC frameshift mutation in a gene from a cancer cell. The peptides are
CC characterised in that they:
CC (i) are at least 8 amino acids long and a fragment of a mutant protein
CC arising from a frameshift mutation in a gene of a cancer cell;
CC (ii) consist of at least one amino acid of the mutant part of a protein
CC sequence encoded by the gene;
CC (iii) comprise 0-10 amino acid from the carboxyl terminus of the normal
CC part of the protein sequence preceding the amino terminus of the mutant
CC sequence and may further extend to the carboxyl terminus of the mutant
CC part of the protein as determined by a new stop codon generated by the
CC frameshift mutation; and
CC (iv) induce, either in their full lengths or after processing by an
CC antigen presenting cell (APC), T cell responses.
CC The genes that the peptides are derived from, are characterised as
CC susceptible to frameshift mutation by having a mono nucleoside base
CC repeat sequence of at least 5 residues, or a di-nucleoside base repeat
CC sequence of at least 4 di-nucleoside base units. The peptides are
CC created by the addition or deletion of 1 or 2 nucleoside base residues
CC from the repeat sequence. The novel peptides can elicit T cell responses
CC and toxicity against tumours and cancer cells carrying genes with
CC frameshift mutations. The novel peptides and DNA sequences can be used
CC for the preparation of a composition for the treatment or prophylaxis of
CC cancer.
XX
SQ Sequence 9 AA;

Query Match 39.1%; Score 9; DB 21; Length 9;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 RLSSCVPVA 12
|||||||||
Db 1 RLSSCVPVA 9

RESULT 21
AAY66123
ID AAY66123 standard; Peptide; 9 AA.
XX
AC AAY66123;
XX
DT 10-FEB-2000 (first entry)
XX
DE Frameshift mutated gene peptide 3.
XX
KW Human; frameshift mutant; T cell response; tumour; treatment; cancer;
KW mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9958552-A2.
XX
PD 18-NOV-1999.
XX
PF 03-MAY-1999; 99WO-NO00143.
XX
PR 08-MAY-1998; 98NO-0002097.
XX
PA (NHYD) NORSK HYDRO AS.
XX
PI Gaudernack G, Eriksen JA, Moller M, Gjertsen MK, Saeterdal I;
XX

DR WPI; 2000-039064/03.
XX
PT New peptides derived from genes with frameshift mutations, used to
PT develop products for the treatment and prophylaxis of cancers -
XX
PS Claim 13; Page 161; 166pp; English.
XX
CC Peptides AAY65684-Y66142 are fragments of mutant proteins arising from a
CC frameshift mutation in a gene from a cancer cell. The peptides are
CC characterised in that they:
CC (i) are at least 8 amino acids long and a fragment of a mutant protein
CC arising from a frameshift mutation in a gene of a cancer cell;
CC (ii) consist of at least one amino acid of the mutant part of a protein
CC sequence encoded by the gene;
CC (iii) comprise 0-10 amino acid from the carboxyl terminus of the normal
CC part of the protein sequence preceding the amino terminus of the mutant
CC sequence and may further extend to the carboxyl terminus of the mutant
CC part of the protein as determined by a new stop codon generated by the
CC frameshift mutation; and
CC (iv) induce, either in their full lengths or after processing by an
CC antigen presenting cell (APC), T cell responses.
CC The genes that the peptides are derived from, are characterised as
CC susceptible to frameshift mutation by having a mono nucleoside base
CC repeat sequence of at least 5 residues, or a di-nucleoside base repeat
CC sequence of at least 4 di-nucleoside base units. The peptides are
CC created by the addition or deletion of 1 or 2 nucleoside base residues
CC from the repeat sequence. The novel peptides can elicit T cell responses
CC and toxicity against tumours and cancer cells carrying genes with
CC frameshift mutations. The novel peptides and DNA sequences can be used
CC for the preparation of a composition for the treatment or prophylaxis of
CC cancer.
XX
SQ Sequence 9 AA;

Query Match 39.1%; Score 9; DB 21; Length 9;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 VRLSSCVPV 11
|||||||||
Db 1 VRLSSCVPV 9

RESULT 22
AAY66124
ID AAY66124 standard; Peptide; 9 AA.
XX
AC AAY66124;
XX
DT 10-FEB-2000 (first entry)
XX
DE Frameshift mutated gene peptide 4.
XX
KW Human; frameshift mutant; T cell response; tumour; treatment; cancer;
KW mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9958552-A2.
XX
PD 18-NOV-1999.
XX
PF 03-MAY-1999; 99WO-NO00143.
XX
PR 08-MAY-1998; 98NO-0002097.
XX
PA (NHYD) NORSK HYDRO AS.
XX
PI Gaudernack G, Eriksen JA, Moller M, Gjertsen MK, Saeterdal I;
XX
DR WPI; 2000-039064/03.
XX
PT New peptides derived from genes with frameshift mutations, used to
PT develop products for the treatment and prophylaxis of cancers -
XX
PS Claim 13; Page 161; 166pp; English.
XX
CC Peptides AAY65684-Y66142 are fragments of mutant proteins arising from a
CC frameshift mutation in a gene from a cancer cell. The peptides are
CC characterised in that they:
CC (i) are at least 8 amino acids long and a fragment of a mutant protein
CC arising from a frameshift mutation in a gene of a cancer cell;
CC (ii) consist of at least one amino acid of the mutant part of a protein
CC sequence encoded by the gene;
CC (iii) comprise 0-10 amino acid from the carboxyl terminus of the normal
CC part of the protein sequence preceding the amino terminus of the mutant
CC sequence and may further extend to the carboxyl terminus of the mutant
CC part of the protein as determined by a new stop codon generated by the
CC frameshift mutation; and
CC (iv) induce, either in their full lengths or after processing by an
CC antigen presenting cell (APC), T cell responses.
CC The genes that the peptides are derived from, are characterised as
CC susceptible to frameshift mutation by having a mono nucleoside base
CC repeat sequence of at least 5 residues, or a di-nucleoside base repeat
CC sequence of at least 4 di-nucleoside base units. The peptides are
CC created by the addition or deletion of 1 or 2 nucleoside base residues
CC from the repeat sequence. The novel peptides can elicit T cell responses
CC and toxicity against tumours and cancer cells carrying genes with
CC frameshift mutations. The novel peptides and DNA sequences can be used
CC for the preparation of a composition for the treatment or prophylaxis of
CC cancer.
XX
SQ Sequence 9 AA;

Query Match 39.1%; Score 9; DB 21; Length 9;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LVRLSSCVP 10
|||||||||
Db 1 LVRLSSCVP 9

RESULT 23
AAY66125
ID AAY66125 standard; Peptide; 9 AA.
XX
AC AAY66125;
XX
DT 10-FEB-2000 (first entry)
XX
DE Frameshift mutated gene peptide 5.
XX
KW Human; frameshift mutant; T cell response; tumour; treatment; cancer;
KW mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9958552-A2.
XX
PD 18-NOV-1999.
XX
PF 03-MAY-1999; 99WO-NO00143.
XX
PR 08-MAY-1998; 98NO-0002097.
XX
PA (NHYD) NORSK HYDRO AS.
XX
PI Gaudernack G, Eriksen JA, Moller M, Gjertsen MK, Saeterdal I;
XX
DR WPI; 2000-039064/03.
XX
PT New peptides derived from genes with frameshift mutations, used to
PT develop products for the treatment and prophylaxis of cancers -

```
XX
PS   Claim 13; Page 161; 166pp; English.
XX
CC   Peptides AAY65684-Y66142 are fragments of mutant proteins arising from a
CC   frameshift mutation in a gene from a cancer cell. The peptides are
CC   characterised in that they:
CC   (i) are at least 8 amino acids long and a fragment of a mutant protein
CC   arising from a frameshift mutation in a gene of a cancer cell;
CC   (ii) consist of at least one amino acid of the mutant part of a protein
CC   sequence encoded by the gene;
CC   (iii) comprise 0-10 amino acid from the carboxyl terminus of the normal
CC   part of the protein sequence preceding the amino terminus of the mutant
CC   sequence and may further extend to the carboxyl terminus of the mutant
CC   part of the protein as determined by a new stop codon generated by the
CC   frameshift mutation; and
CC   (iv) induce, either in their full lengths or after processing by an
CC   antigen presenting cell (APC), T cell responses.
CC   The genes that the peptides are derived from, are characterised as
CC   susceptible to frameshift mutation by having a mono nucleoside base
CC   repeat sequence of at least 5 residues, or a di-nucleoside base repeat
CC   sequence of at least 4 di-nucleoside base units. The peptides are
CC   created by the addition or deletion of 1 or 2 nucleoside base residues
CC   from the repeat sequence. The novel peptides can elicit T cell responses
CC   and toxicity against tumours and cancer cells carrying genes with
CC   frameshift mutations. The novel peptides and DNA sequences can be used
CC   for the preparation of a composition for the treatment or prophylaxis of
CC   cancer.
XX
SQ   Sequence   9 AA;

  Query Match             39.1%;  Score 9;  DB 21;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 7.8e+05;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 SCVPVALMS 15
           |||||||||
Db       1 SCVPVALMS 9


RESULT 24
AAY66126
ID   AAY66126 standard; Peptide; 9 AA.
XX
AC   AAY66126;
XX
DT   10-FEB-2000  (first entry)
XX
DE   Frameshift mutated gene peptide 6.
XX
KW   Human; frameshift mutant; T cell response; tumour; treatment; cancer;
KW   mutein.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9958552-A2.
XX
PD   18-NOV-1999.
XX
PF   03-MAY-1999;   99WO-NO00143.
XX
PR   08-MAY-1998;   98NO-0002097.
XX
PA   (NHYD ) NORSK HYDRO AS.
XX
PI   Gaudernack G,  Eriksen JA,  Moller M,  Gjertsen MK,  Saeterdal I;
XX
DR   WPI; 2000-039064/03.
XX
PT   New peptides derived from genes with frameshift mutations, used to
PT   develop products for the treatment and prophylaxis of cancers -
XX
PS   Claim 13; Page 162; 166pp; English.
XX
CC   Peptides AAY65684-Y66142 are fragments of mutant proteins arising from a
CC   frameshift mutation in a gene from a cancer cell. The peptides are
CC   characterised in that they:
CC   (i) are at least 8 amino acids long and a fragment of a mutant protein
CC   arising from a frameshift mutation in a gene of a cancer cell;
CC   (ii) consist of at least one amino acid of the mutant part of a protein
CC   sequence encoded by the gene;
CC   (iii) comprise 0-10 amino acid from the carboxyl terminus of the normal
CC   part of the protein sequence preceding the amino terminus of the mutant
CC   sequence and may further extend to the carboxyl terminus of the mutant
CC   part of the protein as determined by a new stop codon generated by the
CC   frameshift mutation; and
CC   (iv) induce, either in their full lengths or after processing by an
CC   antigen presenting cell (APC), T cell responses.
CC   The genes that the peptides are derived from, are characterised as
CC   susceptible to frameshift mutation by having a mono nucleoside base
CC   repeat sequence of at least 5 residues, or a di-nucleoside base repeat
CC   sequence of at least 4 di-nucleoside base units. The peptides are
CC   created by the addition or deletion of 1 or 2 nucleoside base residues
CC   from the repeat sequence. The novel peptides can elicit T cell responses
CC   and toxicity against tumours and cancer cells carrying genes with
CC   frameshift mutations. The novel peptides and DNA sequences can be used
CC   for the preparation of a composition for the treatment or prophylaxis of
CC   cancer.
XX
SQ   Sequence   9 AA;

  Query Match             39.1%;  Score 9;  DB 21;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 7.8e+05;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 SSCVPVALM 14
           |||||||||
Db       1 SSCVPVALM 9


RESULT 25
AAY66127
ID   AAY66127 standard; Peptide; 9 AA.
XX
AC   AAY66127;
XX
DT   10-FEB-2000  (first entry)
XX
DE   Frameshift mutated gene peptide 7.
XX
KW   Human; frameshift mutant; T cell response; tumour; treatment; cancer;
KW   mutein.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9958552-A2.
XX
PD   18-NOV-1999.
XX
PF   03-MAY-1999;   99WO-NO00143.
XX
PR   08-MAY-1998;   98NO-0002097.
XX
PA   (NHYD ) NORSK HYDRO AS.
XX
PI   Gaudernack G,  Eriksen JA,  Moller M,  Gjertsen MK,  Saeterdal I;
XX
DR   WPI; 2000-039064/03.
XX
PT   New peptides derived from genes with frameshift mutations, used to
PT   develop products for the treatment and prophylaxis of cancers -
XX
PS   Claim 13; Page 162; 166pp; English.
XX
CC   Peptides AAY65684-Y66142 are fragments of mutant proteins arising from a
```

CC frameshift mutation in a gene from a cancer cell. The peptides are
CC characterised in that they:
CC (i) are at least 8 amino acids long and a fragment of a mutant protein
CC arising from a frameshift mutation in a gene of a cancer cell;
CC (ii) consist of at least one amino acid of the mutant part of a protein
CC sequence encoded by the gene;
CC (iii) comprise 0-10 amino acid from the carboxyl terminus of the normal
CC part of the protein sequence preceding the amino terminus of the mutant
CC sequence and may further extend to the carboxyl terminus of the mutant
CC part of the protein as determined by a new stop codon generated by the
CC frameshift mutation; and
CC (iv) induce, either in their full lengths or after processing by an
CC antigen presenting cell (APC), T cell responses.
CC The genes that the peptides are derived from, are characterised as
CC susceptible to frameshift mutation by having a mono nucleoside base
CC repeat sequence of at least 5 residues, or a di-nucleoside base repeat
CC sequence of at least 4 di-nucleoside base units. The peptides are
CC created by the addition or deletion of 1 or 2 nucleoside base residues
CC from the repeat sequence. The novel peptides can elicit T cell responses
CC and toxicity against tumours and cancer cells carrying genes with
CC frameshift mutations. The novel peptides and DNA sequences can be used
CC for the preparation of a composition for the treatment or prophylaxis of
CC cancer.
XX
SQ Sequence 9 AA;

Query Match 39.1%; Score 9; DB 21; Length 9;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LSSCVPVAL 13
|||||||||
Db 1 LSSCVPVAL 9

RESULT 26
AAY66128
ID AAY66128 standard; Peptide; 9 AA.
XX
AC AAY66128;
XX
DT 10-FEB-2000 (first entry)
XX
DE Frameshift mutated gene peptide 8.
XX
KW Human; frameshift mutant; T cell response; tumour; treatment; cancer;
KW mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9958552-A2.
XX
PD 18-NOV-1999.
XX
PF 03-MAY-1999; 99WO-NO00143.
XX
PR 08-MAY-1998; 98NO-0002097.
XX
PA (NHYD) NORSK HYDRO AS.
XX
PI Gaudernack G, Eriksen JA, Moller M, Gjertsen MK, Saeterdal I;
XX
DR WPI; 2000-039064/03.
XX
PT New peptides derived from genes with frameshift mutations, used to
PT develop products for the treatment and prophylaxis of cancers -
XX
PS Claim 13; Page 162; 166pp; English.
XX
CC Peptides AAY65684-Y66142 are fragments of mutant proteins arising from a
CC frameshift mutation in a gene from a cancer cell. The peptides are
CC characterised in that they:
CC (i) are at least 8 amino acids long and a fragment of a mutant protein
CC arising from a frameshift mutation in a gene of a cancer cell;
CC (ii) consist of at least one amino acid of the mutant part of a protein
CC sequence encoded by the gene;
CC (iii) comprise 0-10 amino acid from the carboxyl terminus of the normal
CC part of the protein sequence preceding the amino terminus of the mutant
CC sequence and may further extend to the carboxyl terminus of the mutant
CC part of the protein as determined by a new stop codon generated by the
CC frameshift mutation; and
CC (iv) induce, either in their full lengths or after processing by an
CC antigen presenting cell (APC), T cell responses.
CC The genes that the peptides are derived from, are characterised as
CC susceptible to frameshift mutation by having a mono nucleoside base
CC repeat sequence of at least 5 residues, or a di-nucleoside base repeat
CC sequence of at least 4 di-nucleoside base units. The peptides are
CC created by the addition or deletion of 1 or 2 nucleoside base residues
CC from the repeat sequence. The novel peptides can elicit T cell responses
CC and toxicity against tumours and cancer cells carrying genes with
CC frameshift mutations. The novel peptides and DNA sequences can be used
CC for the preparation of a composition for the treatment or prophylaxis of
CC cancer.
XX
SQ Sequence 9 AA;

Query Match 39.1%; Score 9; DB 21; Length 9;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 VPVALMSAM 17
|||||||||
Db 1 VPVALMSAM 9

RESULT 27
AAY66129
ID AAY66129 standard; Peptide; 9 AA.
XX
AC AAY66129;
XX
DT 10-FEB-2000 (first entry)
XX
DE Frameshift mutated gene peptide 9.
XX
KW Human; frameshift mutant; T cell response; tumour; treatment; cancer;
KW mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9958552-A2.
XX
PD 18-NOV-1999.
XX
PF 03-MAY-1999; 99WO-NO00143.
XX
PR 08-MAY-1998; 98NO-0002097.
XX
PA (NHYD) NORSK HYDRO AS.
XX
PI Gaudernack G, Eriksen JA, Moller M, Gjertsen MK, Saeterdal I;
XX
DR WPI; 2000-039064/03.
XX
PT New peptides derived from genes with frameshift mutations, used to
PT develop products for the treatment and prophylaxis of cancers -
XX
PS Claim 13; Page 162; 166pp; English.
XX
CC Peptides AAY65684-Y66142 are fragments of mutant proteins arising from a
CC frameshift mutation in a gene from a cancer cell. The peptides are
CC characterised in that they:
CC (i) are at least 8 amino acids long and a fragment of a mutant protein
CC arising from a frameshift mutation in a gene of a cancer cell;

CC (ii) consist of at least one amino acid of the mutant part of a protein
CC sequence encoded by the gene;
CC (iii) comprise 0-10 amino acid from the carboxyl terminus of the normal
CC part of the protein sequence preceding the amino terminus of the mutant
CC sequence and may further extend to the carboxyl terminus of the mutant
CC part of the protein as determined by a new stop codon generated by the
CC frameshift mutation; and
CC (iv) induce, either in their full lengths or after processing by an
CC antigen presenting cell (APC), T cell responses.
CC The genes that the peptides are derived from, are characterised as
CC susceptible to frameshift mutation by having a mono nucleoside base
CC repeat sequence of at least 5 residues, or a di-nucleoside base repeat
CC sequence of at least 4 di-nucleoside base units. The peptides are
CC created by the addition or deletion of 1 or 2 nucleoside base residues
CC from the repeat sequence. The novel peptides can elicit T cell responses
CC and toxicity against tumours and cancer cells carrying genes with
CC frameshift mutations. The novel peptides and DNA sequences can be used
CC for the preparation of a composition for the treatment or prophylaxis of
CC cancer.
XX
SQ Sequence 9 AA;

Query Match 39.1%; Score 9; DB 21; Length 9;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 CVPVALMSA 16
|||||||||
Db 1 CVPVALMSA 9

RESULT 28
AAY54025
ID AAY54025 standard; peptide; 10 AA.
XX
AC AAY54025;
XX
DT 27-MAR-2000 (first entry)
XX
DE Peptide capable of binding MHCI glycoprotein HLA-A3 and HLA-A11.
XX
KW Class I major histocompatibility glycoprotein complex; MHCI;
KW mutant RII receptor; transforming growth factor-beta; TGF-beta;
KW tumour; vaccine; gastric cancer; colon cancer; gene therapy.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN FR2779432-A1.
XX
PD 10-DEC-1999.
XX
PF 08-JUN-1998; 98FR-0007322.
XX
PR 08-JUN-1998; 98FR-0007322.
XX
PA (TRGE) TRANSGENE SA.
XX
DR WPI; 2000-074958/07.
DR N-PSDB; AAZ37063.
XX
PT New nucleic acid sequences, useful for production of medicament for
PT diagnosing, preventing and/or treating gastric or colon cancers -
XX
PS Claim 2; Page 21; 41pp; French.
XX
CC The present sequence represents a peptide which is capable of fixing
CC itself on the glycoproteinS HLA-A3 and HLA-A11 of the class I major
CC histocompatibility glycoprotein complex (MHCI). The specification
CC describes peptides which atatch themselves to at least one MHC1
CC glycoprotien, and which do not comprise the sequence given in AAY54017.
CC The peptides are derived from a mutant RII receptor of transforming
CC growth factor-beta (TGF-beta). The presence of the mutant receptor
CC leads to inactivation of TGF-beta, and contributes to the development
CC of tumours. Especially, the mutation comprises the addition or deletion
CC of an adenine between positions 709-718. The peptides, or nucleic
CC acids encoding them, are useful for the production of a medicament
CC (either preventative, therapeutic or as a vaccine) for treating gastric
CC cancers or cancers of the colon by gene therapy or the peptide may
CC be used as a diagnostic, prophylactic and/or therapeutic composition
CC for the detection, prevention or treatment of gastric or colon cancers.
XX
SQ Sequence 10 AA;

Query Match 39.1%; Score 9; DB 21; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.018;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 SAMTTSSSQ 23
|||||||||
Db 1 SAMTTSSSQ 9

RESULT 29
AAY54029
ID AAY54029 standard; peptide; 10 AA.
XX
AC AAY54029;
XX
DT 27-MAR-2000 (first entry)
XX
DE Peptide which is capable of binding MHCI glycoprotein HLA-B8.
XX
KW Class I major histocompatibility glycoprotein complex; MHCI;
KW mutant RII receptor; transforming growth factor-beta; TGF-beta;
KW tumour; vaccine; gastric cancer; colon cancer; gene therapy.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN FR2779432-A1.
XX
PD 10-DEC-1999.
XX
PF 08-JUN-1998; 98FR-0007322.
XX
PR 08-JUN-1998; 98FR-0007322.
XX
PA (TRGE) TRANSGENE SA.
XX
DR WPI; 2000-074958/07.
DR N-PSDB; AAZ37067.
XX
PT New nucleic acid sequences, useful for production of medicament for
PT diagnosing, preventing and/or treating gastric or colon cancers -
XX
PS Claim 2; Page 22; 41pp; French.
XX
CC The present sequence represents a peptide which is capable of fixing
CC itself on the glycoprotein HLA-B8 of the class I major
CC histocompatibility glycoprotein complex (MHCI). The specification
CC describes peptides which atatch themselves to at least one MHC1
CC glycoprotien, and which do not comprise the sequence given in AAY54017.
CC The peptides are derived from a mutant RII receptor of transforming
CC growth factor-beta (TGF-beta). The presence of the mutant receptor leads
CC to inactivation of TGF-beta, and contributes to the development of
CC tumours. Especially, the mutation comprises the addition or deletion of
CC an adenine between positions 709-718. The peptides, or nucleic acids
CC encoding them, are useful for the production of a medicament (either
CC preventative, therapeutic or as a vaccine) for treating gastric cancers
CC or cancers of the colon by gene therapy or the peptide may be used as a
CC diagnostic, prophylactic and/or therapeutic composition for the
CC detection, prevention or treatment of gastric or colon cancers.
XX
SQ Sequence 10 AA;

```
  Query Match             39.1%;  Score 9;  DB 21;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 0.018;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SLVRLSSCV 9
           |||||||||
Db       2 SLVRLSSCV 10


RESULT 30
AAY65698
ID   AAY65698 standard; Peptide; 19 AA.
XX
AC   AAY65698;
XX
DT   10-FEB-2000  (first entry)
XX
DE   TGF beta RII mutant peptide 3.
XX
KW   Human; frameshift mutant; T cell response; tumour; treatment; cancer;
KW   mutein.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9958552-A2.
XX
PD   18-NOV-1999.
XX
PF   03-MAY-1999;   99WO-NO00143.
XX
PR   08-MAY-1998;   98NO-0002097.
XX
PA   (NHYD ) NORSK HYDRO AS.
XX
PI   Gaudernack G,  Eriksen JA,  Moller M,  Gjertsen MK,  Saeterdal I;
XX
DR   WPI; 2000-039064/03.
XX
PT   New peptides derived from genes with frameshift mutations, used to
PT   develop products for the treatment and prophylaxis of cancers -
XX
PS   Claim 12; Page 20; 166pp; English.
XX
CC   Peptides AAY65684-Y66142 are fragments of mutant proteins arising from a
CC   frameshift mutation in a gene from a cancer cell. The peptides are
CC   characterised in that they:
CC   (i) are at least 8 amino acids long and a fragment of a mutant protein
CC   arising from a frameshift mutation in a gene of a cancer cell;
CC   (ii) consist of at least one amino acid of the mutant part of a protein
CC   sequence encoded by the gene;
CC   (iii) comprise 0-10 amino acid from the carboxyl terminus of the normal
CC   part of the protein sequence preceding the amino terminus of the mutant
CC   sequence and may further extend to the carboxyl terminus of the mutant
CC   part of the protein as determined by a new stop codon generated by the
CC   frameshift mutation; and
CC   (iv) induce, either in their full lengths or after processing by an
CC   antigen presenting cell (APC), T cell responses.
CC   The genes that the peptides are derived from, are characterised as
CC   susceptible to frameshift mutation by having a mono nucleoside base
CC   repeat sequence of at least 5 residues, or a di-nucleoside base repeat
CC   sequence of at least 4 di-nucleoside base units. The peptides are
CC   created by the addition or deletion of 1 or 2 nucleoside base residues
CC   from the repeat sequence. The novel peptides can elicit T cell responses
CC   and toxicity against tumours and cancer cells carrying genes with
CC   frameshift mutations. The novel peptides and DNA sequences can be used
CC   for the preparation of a composition for the treatment or prophylaxis of
CC   cancer.
XX
SQ   Sequence   19 AA;

  Query Match             39.1%;  Score 9;  DB 21;  Length 19;
  Best Local Similarity   100.0%;  Pred. No. 0.033;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SLVRLSSCV 9
           |||||||||
Db      11 SLVRLSSCV 19


RESULT 31
AAY54020
ID   AAY54020 standard; peptide; 8 AA.
XX
AC   AAY54020;
XX
DT   27-MAR-2000  (first entry)
XX
DE   Peptide which is capable of binding MHCI glycoprotein HLA-A2.
XX
KW   Class I major histocompatibility glycoprotein complex; MHCI;
KW   mutant RII receptor; transforming growth factor-beta; TGF-beta;
KW   tumour; vaccine; gastric cancer; colon cancer; gene therapy.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   FR2779432-A1.
XX
PD   10-DEC-1999.
XX
PF   08-JUN-1998;   98FR-0007322.
XX
PR   08-JUN-1998;   98FR-0007322.
XX
PA   (TRGE ) TRANSGENE SA.
XX
DR   WPI; 2000-074958/07.
DR   N-PSDB; AAZ37058.
XX
PT   New nucleic acid sequences, useful for production of medicament for
PT   diagnosing, preventing and/or treating gastric or colon cancers -
XX
PS   Claim 2; Page 20; 41pp; French.
XX
CC   The present sequence represents a peptide which is capable of fixing
CC   itself on the glycoprotein HLA-A2 of the class I major
CC   histocompatibility glycoprotein complex (MHCI). The specification
CC   describes peptides which atatch themselves to at least one MHCl
CC   glycoprotien, and which do not comprise the sequence given in AAY54017.
CC   The peptides are derived from a mutant RII receptor of transforming
CC   growth factor-beta (TGF-beta). The presence of the mutant receptor leads
CC   to inactivation of TGF-beta, and contributes to the development of
CC   tumours. Especially, the mutation comprises the addition or deletion of
CC   an adenine between positions 709-718. The peptides, or nucleic acids
CC   encoding them, are useful for the production of a medicament (either
CC   preventative, therapeutic or as a vaccine) for treating gastric cancers
CC   or cancers of the colon by gene therapy or the peptide may be used as a
CC   diagnostic, prophylactic and/or therapeutic composition for the
CC   detection, prevention or treatment of gastric or colon cancers.
XX
SQ   Sequence   8 AA;

  Query Match             34.8%;  Score 8;  DB 21;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 7.8e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 RLSSCVPV 11
           ||||||||
Db       1 RLSSCVPV 8


RESULT 32
AAY54034
ID   AAY54034 standard; peptide; 8 AA.
XX
```

```
AC   AAY54034;
XX
DT   27-MAR-2000  (first entry)
XX
DE   Peptide which is capable of binding MHCI glycoprotein HLA-B7.
XX
KW   Class I major histocompatibility glycoprotein complex; MHCI;
KW   mutant RII receptor; transforming growth factor-beta; TGF-beta;
KW   tumour; vaccine; gastric cancer; colon cancer; gene therapy.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   FR2779432-A1.
XX
PD   10-DEC-1999.
XX
PF   08-JUN-1998;   98FR-0007322.
XX
PR   08-JUN-1998;   98FR-0007322.
XX
PA   (TRGE ) TRANSGENE SA.
XX
DR   WPI; 2000-074958/07.
DR   N-PSDB; AAZ37072.
XX
PT   New nucleic acid sequences, useful for production of medicament for
PT   diagnosing, preventing and/or treating gastric or colon cancers -
XX
PS   Claim 2; Page 24; 41pp; French.
XX
CC   The present sequence represents a peptide which is capable of fixing
CC   itself on the glycoprotein HLA-B7 of the class I major
CC   histocompatibility glycoprotein complex (MHCI). The specification
CC   describes peptides which atatch themselves to at least one MHC1
CC   glycoprotien, and which do not comprise the sequence given in AAY54017.
CC   The peptides are derived from a mutant RII receptor of transforming
CC   growth factor-beta (TGF-beta). The presence of the mutant receptor leads
CC   to inactivation of TGF-beta, and contributes to the development of
CC   tumours. Especially, the mutation comprises the addition or deletion of
CC   an adenine between positions 709-718. The peptides, or nucleic acids
CC   encoding them, are useful for the production of a medicament (either
CC   preventative, therapeutic or as a vaccine) for treating gastric cancers
CC   or cancers of the colon by gene therapy or the peptide may be used as a
CC   diagnostic, prophylactic and/or therapeutic composition for the
CC   detection, prevention or treatment of gastric or colon cancers.
XX
SQ   Sequence   8 AA;

  Query Match             34.8%;  Score 8;  DB 21;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 7.8e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 VPVALMSA 16
            ||||||||
Db       1 VPVALMSA 8


RESULT 33
AAY54024
ID   AAY54024 standard; peptide; 9 AA.
XX
AC   AAY54024;
XX
DT   27-MAR-2000  (first entry)
XX
DE   Peptide capable of binding MHCI glycoprotein HLA-A3 and HLA-A11.
XX
KW   Class I major histocompatibility glycoprotein complex; MHCI;
KW   mutant RII receptor; transforming growth factor-beta; TGF-beta;
KW   tumour; vaccine; gastric cancer; colon cancer; gene therapy.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   FR2779432-A1.
XX
PD   10-DEC-1999.
XX
PF   08-JUN-1998;   98FR-0007322.
XX
PR   08-JUN-1998;   98FR-0007322.
XX
PA   (TRGE ) TRANSGENE SA.
XX
DR   WPI; 2000-074958/07.
DR   N-PSDB; AAZ37062.
XX
PT   New nucleic acid sequences, useful for production of medicament for
PT   diagnosing, preventing and/or treating gastric or colon cancers -
XX
PS   Claim 2; Page 21; 41pp; French.
XX
CC   The present sequence represents a peptide which is capable of fixing
CC   itself on the glycoproteinS HLA-A3 and HLA-A11 of the class I major
CC   histocompatibility glycoprotein complex (MHCI). The specification
CC   describes peptides which atatch themselves to at least one MHC1
CC   glycoprotien, and which do not comprise the sequence given in AAY54017.
CC   The peptides are derived from a mutant RII receptor of transforming
CC   growth factor-beta (TGF-beta). The presence of the mutant receptor
CC   leads to inactivation of TGF-beta, and contributes to the development
CC   of tumours. Especially, the mutation comprises the addition or deletion
CC   of an adenine between positions 709-718. The peptides, or nucleic
CC   acids encoding them, are useful for the production of a medicament
CC   (either preventative, therapeutic or as a vaccine) for treating gastric
CC   cancers or cancers of the colon by gene therapy or the peptide may
CC   be used as a diagnostic, prophylactic and/or therapeutic composition
CC   for the detection, prevention or treatment of gastric or colon cancers.
XX
SQ   Sequence   9 AA;

  Query Match             34.8%;  Score 8;  DB 21;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 7.8e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      16 AMTTSSSQ 23
           ||||||||
Db       1 AMTTSSSQ 8


RESULT 34
AAY65704
ID   AAY65704 standard; Peptide; 19 AA.
XX
AC   AAY65704;
XX
DT   10-FEB-2000  (first entry)
XX
DE   TGF beta RII mutant peptide 9.
XX
KW   Human; frameshift mutant; T cell response; tumour; treatment; cancer;
KW   mutein.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9958552-A2.
XX
PD   18-NOV-1999.
XX
PF   03-MAY-1999;   99WO-NO00143.
XX
PR   08-MAY-1998;   98NO-0002097.
XX
PA   (NHYD ) NORSK HYDRO AS.
XX
```

PI Gaudernack G, Eriksen JA, Moller M, Gjertsen MK, Saeterdal I;
XX
DR WPI; 2000-039064/03.
XX
PT New peptides derived from genes with frameshift mutations, used to
PT develop products for the treatment and prophylaxis of cancers -
XX
PS Claim 12; Page 20; 166pp; English.
XX
CC Peptides AAY65684-Y66142 are fragments of mutant proteins arising from a
CC frameshift mutation in a gene from a cancer cell. The peptides are
CC characterised in that they:
CC (i) are at least 8 amino acids long and a fragment of a mutant protein
CC arising from a frameshift mutation in a gene of a cancer cell;
CC (ii) consist of at least one amino acid of the mutant part of a protein
CC sequence encoded by the gene;
CC (iii) comprise 0-10 amino acid from the carboxyl terminus of the normal
CC part of the protein sequence preceding the amino terminus of the mutant
CC sequence and may further extend to the carboxyl terminus of the mutant
CC part of the protein as determined by a new stop codon generated by the
CC frameshift mutation; and
CC (iv) induce, either in their full lengths or after processing by an
CC antigen presenting cell (APC), T cell responses.
CC The genes that the peptides are derived from, are characterised as
CC susceptible to frameshift mutation by having a mono nucleoside base
CC repeat sequence of at least 5 residues, or a di-nucleoside base repeat
CC sequence of at least 4 di-nucleoside base units. The peptides are
CC created by the addition or deletion of 1 or 2 nucleoside base residues
CC from the repeat sequence. The novel peptides can elicit T cell responses
CC and toxicity against tumours and cancer cells carrying genes with
CC frameshift mutations. The novel peptides and DNA sequences can be used
CC for the preparation of a composition for the treatment or prophylaxis of
CC cancer.
XX
SQ Sequence 19 AA;

Query Match 34.8%; Score 8; DB 21; Length 19;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 AMTTSSSQ 23
||||||||
Db 1 AMTTSSSQ 8

RESULT 35
AAY12053
ID AAY12053 standard; Protein; 35 AA.
XX
AC AAY12053;
XX
DT 18-JUN-1999 (first entry)
XX
XX
DE Human 5' EST secreted protein SEQ ID NO: 366.
XX
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition.
XX
OS Homo sapiens.
XX
PN WO9906554-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-IB01238.
XX
PR 01-AUG-1997; 97US-0905134.
XX
PA (GEST) GENSET.
XX
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
XX
DR WPI; 1999-153784/13.
DR N-PSDB; AAX40886.
XX
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from kidney, fetal kidney, dystrophic
PT muscle, muscle and heart tissue
XX
PS Claim 34; Page 492; 622pp; English.
XX
CC AAX40826 to AAX41093 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins, and encode the proteins given in AAY01602 and
CC AAY11994 to AAY12260, respectively. The proteins given represent the
CC signal peptide and an N-terminal fragment of a secreted protein. The
CC nucleic acid sequences can be used for producing secreted human gene
CC products. They can also be used to develop products for diagnosis and
CC therapy. The proteins obtained may have cytokine activity, cell
CC proliferation/differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptide can be used
CC for directing extracellular secretion of a polypeptide or the insertion
CC of a polypeptide into a membrane, or importing a polypeptide into
CC a cell.
XX
SQ Sequence 35 AA;

Query Match 34.8%; Score 8; DB 20; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.56;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 SSCVPVAL 13
||||||||
Db 27 SSCVPVAL 34

part of SQ 9 (17)

Search completed: May 7, 2003, 09:32:23
Job time : 35 secs

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: May 7, 2003, 09:31:46 ; Search time 17 Seconds
(without alignments)
124.505 Million cell updates/sec

Title: US-09-674-973A-17
Perfect score: 23
Sequence: 1 SLVRLSSCVPVALMSAMTTSSSQ 23

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 349150 seqs, 92025710 residues

Word size : 8

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	23	100.0	34	10	US-09-878-905-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-878-905-3
; Sequence 3, Application US/09878905
; Patent No. US20020064786A1
; GENERAL INFORMATION:
; APPLICANT: Markowitz, Sanford D
; APPLICANT: Brattain, Michael G
; APPLICANT: Willson, James K.V.
; TITLE OF INVENTION: CANCER DIAGNOSIS, PROGNOSIS AND THERAPY BASED ON
; TITLE OF INVENTION: MUTATION OF RECEPTOR
; FILE REFERENCE: 062361.0108
; CURRENT APPLICATION NUMBER: US/09/878,905
; CURRENT FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: 08/417,867
; PRIOR FILING DATE: 1995-04-07
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 34
; TYPE: PRT
; ORGANISM: human
US-09-878-905-3

Query Match 100.0%; Score 23; DB 10; Length 34;
Best Local Similarity 100.0%; Pred. No. 4.5e-16;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLVRLSSCVPVALMSAMTTSSSQ 23
||||||||||||||||||||||||
Db 1 SLVRLSSCVPVALMSAMTTSSSQ 23

Search completed: May 7, 2003, 09:34:03
Job time : 17 secs

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: May 7, 2003, 09:31:26 ; Search time 15 Seconds
(without alignments)
147.406 Million cell updates/sec

Title: US-09-674-973A-17
Perfect score: 23
Sequence: 1 SLVRLSSCVPVALMSAMTTSSSQ 23

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 8

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description

No matches found

Search completed: May 7, 2003, 09:33:39
Job time : 15 secs

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: May 7, 2003, 09:30:25 ; Search time 11 Seconds
(without alignments)
86.723 Million cell updates/sec

Title: US-09-674-973A-17
Perfect score: 23
Sequence: 1 SLVRLSSCVPVALMSAMTTSSSQ 23

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 8

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description

No matches found

Search completed: May 7, 2003, 09:32:40
Job time : 11 secs

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: May 7, 2003, 09:31:01 ; Search time 28 Seconds
(without alignments)
169.253 Million cell updates/sec

Title: US-09-674-973A-17
Perfect score: 23
Sequence: 1 SLVRLSSCVPVALMSAMTTSSSQ 23

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 8

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description

No matches found

Search completed: May 7, 2003, 09:33:16
Job time : 28 secs

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: May 7, 2003, 09:28:55 ; Search time 15 Seconds
(without alignments)
147.406 Million cell updates/sec

Title: US-09-674-973A-17
Perfect score: 106
Sequence: 1 SLVRLSSCVPVALMSAMTTSSSQ 23

›ring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	52	49.1	99	2	S40012	fill protein - gar
2	45.5	42.9	286	2	H70812	hypothetical prote
3	45	42.5	102	2	T35134	hypothetical prote
4	43	40.6	227	2	G69762	two-component resp
5	43	40.6	251	2	C72572	hypothetical prote
6	43	40.6	273	2	F86665	ABC transporter pe
7	43	40.6	314	2	H86354	hypothetical prote
8	43	40.6	337	2	E96037	probable ABC trans
9	43	40.6	970	2	F87450	TonB-dependent rec
10	42	39.6	539	2	T46720	hypothetical prote
11	41	38.7	96	2	T04060	probable molybdopt
12	41	38.7	355	1	B69518	GTP-binding protei
13	41	38.7	601	2	S74239	secretogranin II p
14	41	38.7	652	2	T20046	hypothetical prote
15	41	38.7	1101	2	T16840	hypothetical prote
16	41	38.7	2233	2	T28669	surface protein 51
17	40.5	38.2	242	2	AF2543	hypothetical prote
18	40.5	38.2	581	2	AB3588	probable adenine d
19	40	37.7	194	2	A23277	gamma-secalin - ry
20	40	37.7	242	2	E85072	hypothetical prote
21	40	37.7	284	2	G82932	spermidine/putresc
22	40	37.7	303	2	D70955	hypothetical prote
23	40	37.7	377	2	F64245	hypothetical prote
24	40	37.7	543	1	A54116	cytochrome P450 1B
25	40	37.7	1234	2	T30160	hypothetical prote
26	40	37.7	1243	2	JC5615	membrane-associate
27	40	37.7	1367	1	S48478	glucan 1,4-alpha-g
28	40	37.7	1419	2	T30531	agglutinin-like ad
29	39.5	37.3	263	2	S17661	hemolymph 30K prot
30	39.5	37.3	2210	1	RRXPLC	genome polyprotein
31	39	36.8	160	2	JU0342	rhizopuspepsin (EC
32	39	36.8	220	2	AD1207	transcription resp
33	39	36.8	234	2	D83752	two-component resp
34	39	36.8	308	2	I56573	synaptic glycoprot
35	39	36.8	320	2	D86259	protein T12C24.6 [
36	39	36.8	397	2	C95221	hypothetical prote
37	39	36.8	397	2	B98085	hypothetical prote
38	39	36.8	426	2	S64748	mitochondrial oute
39	39	36.8	440	2	AF2359	hypothetical prote
40	39	36.8	512	2	G70662	probable plcB prot
41	39	36.8	519	2	C87350	major facilitator
42	39	36.8	559	2	D90291	permease, multidru
43	39	36.8	560	2	A83155	probable medium-ch
44	39	36.8	645	2	JC5517	Gu/RNA helicase II
45	39	36.8	739	2	A55314	glycine-tRNA ligas

ALIGNMENTS

RESULT 1
S40012
fill protein - garden snapdragon
C;Species: Antirrhinum majus (garden snapdragon)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 01-Dec-2000
C;Accession: S40012; S17699
R;Nacken, W.K.F.
submitted to the EMBL Data Library, January 1991
A;Reference number: S40012
A;Accession: S40012
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-99 <NAC>
A;Cross-references: EMBL:X57296; NID:g406308; PID:g406309
R;Nacken, W.K.F.; Huijser, P.; Beltran, J.P.; Saedler, H.; Sommer, H.
Mol. Gen. Genet. 229, 129-136, 1991
A;Title: Molecular characterization of two stamen-specific genes, tap1 and fil1, that
A;Reference number: S17698; MUID:91375441; PMID:1680216
A;Accession: S17699
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-91,'AN' <NA2>
C;Genetics:
A;Gene: fil1
A;Introns: 92/1

Query Match 49.1%; Score 52; DB 2; Length 99;
Best Local Similarity 45.5%; Pred. No. 0.44;
Matches 10; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

```
Qy       1 SLVRLSSCVPVALMSAMTTSSS 22
           ||  |::| |  :: | || ||
Db      34 SLANLNACAPFVVLGAATTPSS 55
```

RESULT 2
H70812
hypothetical protein Rv0840c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: H70812
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: H70812
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-286 <COL>

```
A;Cross-references: GB:AL022004; GB:AL123456; NID:g3261550; PIDN:CAA17646.1; PID:g291689
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv0840c
C;Superfamily: leucyl aminopeptidase pepL

  Query Match             42.9%;  Score 45.5;  DB 2;  Length 286;
  Best Local Similarity   43.3%;  Pred. No. 13;
  Matches   13;  Conservative    4;  Mismatches    6;  Indels    7;  Gaps    1;

Qy       1 SLVRLSSCVPVALMSAM-------TTSSSQ 23
           ||| | ||: ||  ||:       || |::
Db     133 SLVSLKSCLDVATRSAIDRHEAAGTTHSAE 162


RESULT 3
T35134
hypothetical protein SC4H8.03 SC4H8.03 - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C;Accession: T35134
R;Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, December 1997
A;Reference number: Z21569
A;Accession: T35134
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-102 <HAR>
A;Cross-references: EMBL:AL020958; PIDN:CAA15870.1; GSPDB:GN00070; SCOEDB:SC4H8.03
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SC4H8.03

  Query Match             42.5%;  Score 45;  DB 2;  Length 102;
  Best Local Similarity   42.1%;  Pred. No. 5.7;
  Matches    8;  Conservative    7;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       4 RLSSCVPVALMSAMTTSSS 22
           ||:: :|||||::|:|   :
Db      36 RLAALLPVALLAALTAQQT 54


RESULT 4
G69762
two-component response regulator [YclK] homolog yclJ - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C;Accession: G69762
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
h, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
etter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: G69762
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-227 <KUN>
A;Cross-references: GB:Z99106; GB:AL009126; NID:g2632653; PIDN:CAB12183.1; PID:g2632676
A;Experimental source: strain 168
C;Genetics:
A;Gene: yclJ
C;Superfamily: ompR protein; response regulator homology
C;Keywords: phosphoprotein
F;3-111/Domain: response regulator homology <RRH>
F;51/Binding site: phosphate (Asp) (covalent) #status predicted

  Query Match             40.6%;  Score 43;  DB 2;  Length 227;
  Best Local Similarity   38.1%;  Pred. No. 25;
  Matches    8;  Conservative    6;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       3 VRLSSCVPVALMSAMTTSSSQ 23
           :| :| ||: :::|  | | |
Db      66 IRETSTVPIIMLTAKDTESDQ 86


RESULT 5
C72572
hypothetical protein APE1860 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jun-2000
C;Accession: C72572
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero
A;Reference number: A72450; MUID:99310339; PMID:10382966
A;Accession: C72572
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-251 <KAW>
A;Cross-references: DDBJ:AP000062; NID:g5105244; PIDN:BAA80864.1; PID:d1044650; PID:g
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE1860
C;Superfamily: Aeropyrum pernix hypothetical protein APE1860

  Query Match             40.6%;  Score 43;  DB 2;  Length 251;
  Best Local Similarity   52.9%;  Pred. No. 28;
  Matches    9;  Conservative    4;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 SLVRLSSCVPVALMSAM 17
           ||:||:| ||  ::| |
Db     185 SLLRLASSVPSLILSTM 201


RESULT 6
F86665
ABC transporter permease protein ydcF [imported] - Lactococcus lactis subsp. lactis (
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C;Accession: F86665
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Eh
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: F86665
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-273 <STO>
A;Cross-references: GB:AE005176; PID:g12723192; PIDN:AAK04424.1; GSPDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: ydcF

  Query Match             40.6%;  Score 43;  DB 2;  Length 273;
  Best Local Similarity   36.4%;  Pred. No. 30;
  Matches    8;  Conservative    8;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       2 LVRLSSCVPVALMSAMTTSSSQ 23
           :::  | :|:||:  :||: ||
Db     118 VLKYFSTIPLALLFILTTNPSQ 139


RESULT 7
H86354
```

```
hypothetical protein F16L1.5 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C;Accession: H86354
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: H86354
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-314 <STO>
A;Cross-references: GB:AE005172; NID:g9454528; PIDN:AAF87851.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1
C;Superfamily: Arabidopsis thaliana hypothetical protein F28J12.40

  Query Match             40.6%;  Score 43;  DB 2;  Length 314;
  Best Local Similarity   64.7%;  Pred. No. 34;
  Matches   11;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       5 LSSCVPVALMSAMTTSS 21
           | ||| ||  || | ||
Db     131 LKSCVIVAFRSAGTVSS 147


RESULT 8
E96037
probable ABC transporter permease protein SMb21646 [imported] - Sinorhizobium meliloti (
C;Species: Sinorhizobium meliloti
C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C;Accession: E96037
R;Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A;Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo
A;Reference number: A95842; MUID:21396508; PMID:11481431
A;Accession: E96037
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-337 <KUR>
A;Cross-references: GB:AL591985; PIDN:CAC49965.1; PID:g15141453; GSPDB:GN00167
A;Experimental source: strain 1021, megaplasmid pSymB
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
 L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A;Reference number: A96039; MUID:21368234; PMID:11474104
A;Contents: annotation
C;Genetics:
A;Gene: SMb21646
A;Genome: plasmid
C;Superfamily: oligopeptide permease protein oppB

  Query Match             40.6%;  Score 43;  DB 2;  Length 337;
  Best Local Similarity   44.4%;  Pred. No. 37;
  Matches    8;  Conservative    6;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 LVRLSSCVPVALMSAMTT 19
           |||::| :|| |: :: |
Db       6 LVRIASAIPVLLVLSVVT 23


RESULT 9
F87450
TonB-dependent receptor [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: F87450
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: F87450
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-970 <STO>
A;Cross-references: GB:AE005673; NID:g13423024; PIDN:AAK23602.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC1623

  Query Match             40.6%;  Score 43;  DB 2;  Length 970;
  Best Local Similarity   38.9%;  Pred. No. 1e+02;
  Matches    7;  Conservative    5;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       6 SSCVPVALMSAMTTSSSQ 23
           : |||: | | || : ::
Db     490 AGCVPINLFSPMTAAQAE 507


RESULT 10
T46720
hypothetical protein L4326.09 [imported] - Leishmania major
C;Species: Leishmania major
C;Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 04-Mar-2000
C;Accession: T46720
R;Volckaert, G.; Ivens, A.C.; Lawson, D.; Quail, M.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, December 1999
A;Reference number: Z23137
A;Accession: T46720
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-539 <VOL>
A;Cross-references: EMBL:AL121861; PIDN:CAB58385.1
A;Experimental source: strain Friedlin
C;Genetics:
A;Note: L4326.09
C;Superfamily: Leishmania major hypothetical protein L4326.09

  Query Match             39.6%;  Score 42;  DB 2;  Length 539;
  Best Local Similarity   45.8%;  Pred. No. 82;
  Matches   11;  Conservative    4;  Mismatches    7;  Indels    2;  Gaps    1;

Qy       2 LVRLSSCVP--VALMSAMTTSSSQ 23
           ||| ::|||     ||| :| | :
Db      58 LVRYTACVPAHTGSMSASSTVSDR 81


RESULT 11
T04060
probable molybdopterin synthase small chain F28M11.20 [similarity] - Arabidopsis thal
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 19-Jan-2001
C;Accession: T04060
R;Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.
submitted to the Protein Sequence Database, March 1999
A;Reference number: Z15184
A;Accession: T04060
A;Molecule type: DNA
A;Residues: 1-96 <BEV>
A;Cross-references: EMBL:AL049487
A;Experimental source: cultivar Columbia; BAC clone F28M11
C;Genetics:
A;Map position: 4
A;Note: F28M11.20
```

```
C;Keywords: molybdopterin biosynthesis
F;96/Modified site: 1-thioglycine (Gly) #status predicted

  Query Match             38.7%;  Score 41;  DB 2;  Length 96;
  Best Local Similarity   45.5%;  Pred. No. 23;
  Matches   10;  Conservative    3;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       1 SLVRLSSCVPVALMSAMTTSSS 22
           ||  : ||| :||    || |:
Db      58 SLEEVRSCVVLALNEEYTTDSA 79


RESULT 12
B69518
GTP-binding protein DRG homolog - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 19-Jan-2001
C;Accession: B69518
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A;Reference number: A69250; MUID:98049343; PMID:9389475
A;Accession: B69518
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-355 <KLE>
A;Cross-references: GB:AE000956; GB:AE000782; NID:g2689279; PIDN:AAB89108.1; PID:g264839
C;Superfamily: GTP-binding protein DRG; translation elongation factor Tu homology
C;Keywords: GTP binding; nucleotide binding; P-loop
F;64-183/Domain: translation elongation factor Tu homology <ETU>
F;70-77/Region: nucleotide-binding motif A (P-loop)
F;93-98/Region: GTP binding #status predicted
F;116-119/Region: GTP binding #status predicted
F;245-248/Region: GTP binding #status predicted
F;329-333/Region: GTP binding #status predicted

  Query Match             38.7%;  Score 41;  DB 1;  Length 355;
  Best Local Similarity   50.0%;  Pred. No. 80;
  Matches    9;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       3 VRLSSCVPVALMSAMTTS 20
           ||:|| ||:::  |  ||
Db     187 VRISSTVPLSIDEATITS 204


RESULT 13
S74239
secretogranin II precursor - laughing frog
C;Species: Rana ridibunda (laughing frog)
C;Date: 29-Jan-1998 #sequence_revision 13-Feb-1998 #text_change 15-Oct-1999
C;Accession: S74239; S15867
R;Anouar, Y.; Jegou, S.; Alexandre, D.; Lihrmann, I.; Conlon, J.M.; Vaudry, H.
FEBS Lett. 394, 295-299, 1996
A;Title: Molecular cloning of frog secretogranin II reveals the occurrence of several hi
A;Reference number: S74239; MUID:96427274; PMID:8830661
A;Accession: S74239
A;Molecule type: mRNA
A;Residues: 1-601 <ANO>
A;Cross-references: EMBL:U68757; NID:g1633645; PIDN:AAB17470.1; PID:g1633646
A;Experimental source: pituitary gland
R;Vaudry, H.; Conlon, J.M.
FEBS Lett. 284, 31-33, 1991
A;Title: Identification of a peptide arising from the specific post-translation processi
A;Reference number: S15867; MUID:91285100; PMID:2060624
A;Accession: S15867
A;Molecule type: protein
A;Residues: 183-215 <FEB>
C;Superfamily: secretogranin II
C;Keywords: glycoprotein; pituitary; sulfoprotein
F;1-27/Domain: signal sequence #status predicted <SIG>
F;28-601/Product: secretogranin II #status predicted <MAT>
F;151/Binding site: sulfate (Tyr) (covalent) #status predicted
F;307/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             38.7%;  Score 41;  DB 2;  Length 601;
  Best Local Similarity   52.6%;  Pred. No. 1.3e+02;
  Matches   10;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       5 LSSCVPVALMSAMTTSSSQ 23
           ||||: | |||    :| |
Db      13 LSSCILVILMSFSDAASFQ 31


RESULT 14
T20046
hypothetical protein C49A1.9 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T20046
R;Matthews, L.
submitted to the EMBL Data Library, December 1996
A;Reference number: Z19217
A;Accession: T20046
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-652 <WIL>
A;Cross-references: EMBL:Z83221; PIDN:CAB05706.1; GSPDB:GN00019; CESP:C49A1.9
A;Experimental source: clone C49A1
C;Genetics:
A;Gene: CESP:C49A1.9
A;Map position: 1
A;Introns: 26/1; 70/3; 124/3; 173/3; 213/1; 254/3; 306/2; 335/1; 379/2; 400/3; 427/1;

  Query Match             38.7%;  Score 41;  DB 2;  Length 652;
  Best Local Similarity   45.0%;  Pred. No. 1.4e+02;
  Matches    9;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       4 RLSSCVPVALMSAMTTSSSQ 23
           | |||:| |  ||   |:::
Db     236 RKSSCMPPAKKSASCNSTNE 255


RESULT 15
T16840
hypothetical protein T10E10.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T16840
R;Geisel, C.
submitted to the EMBL Data Library, October 1995
A;Description: The sequence of C. elegans cosmid T10E10.
A;Reference number: Z18588
A;Accession: T16840
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1101 <GEI>
A;Cross-references: EMBL:U39644; NID:g1049339; PID:g1049343; PIDN:AAA80360.1; CESP:T1
A;Experimental source: strain Bristol N2
C;Genetics:
A;Gene: CESP:T10E10.4
A;Introns: 93/2; 152/2; 191/3; 209/2; 283/3; 303/1; 399/3; 421/1; 440/1; 465/1; 547/3

  Query Match             38.7%;  Score 41;  DB 2;  Length 1101;
  Best Local Similarity   36.8%;  Pred. No. 2.3e+02;
  Matches    7;  Conservative    7;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       5 LSSCVPVALMSAMTTSSSQ 23
           |: |||:|: ::  :|: |
Db     361 LNVCVPLAIQNSCDSSTQQ 379


Search completed: May  7, 2003, 09:31:20
```

Job time : 18 secs

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: May 7, 2003; 09:28:05 ; Search time 11 Seconds
(without alignments)
86.723 Million cell updates/sec

Title: US-09-674-973A-17
Perfect score: 106
Sequence: 1 SLVRLSSCVPVALMSAMTTSSSQ 23

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	52	49.1	99	1	FIL1_ANTMA	Q38737 antirrhinum
2	41	38.7	308	1	GSN2_MOUSE	Q9cy27 mus musculu
3	41	38.7	471	1	UDPG_PYRPY	O64459 pyrus pyrif
4	41	38.7	601	1	SG2_RANRI	P30945 rana ridibu
5	40	37.7	154	1	IL2_MIRAN	O62641 mirounga an
6	40	37.7	377	1	Y412_MYCGE	P47652 mycoplasma
7	40	37.7	429	1	CAT2_CLOKL	P38942 clostridium
8	40	37.7	543	1	CP1B_HUMAN	Q16678 homo sapien
9	40	37.7	1367	1	AMYH_YEAST	P08640 saccharomyc
10	40	37.7	1419	1	ALA1_CANAL	O13368 candida alb
11	39.5	37.3	263	1	L302_BOMMO	Q00801 bombyx mori
12	39.5	37.3	2210	1	RRPO_LYCVA	P14240 lymphocytic
13	39	36.8	308	1	GSN2_RAT	Q64232 rattus norv
14	39	36.8	391	1	CAR2_RHINI	P43231 rhizopus ni
15	39	36.8	426	1	MMM1_YEAST	P41800 saccharomyc
16	39	36.8	521	1	PHLB_MYCTU	P95246 mycobacteri
17	39	36.8	651	1	PIA1_HUMAN	O75925 homo sapien
18	39	36.8	651	1	PIA1_MOUSE	O88907 mus musculu
19	39	36.8	701	1	CG1_HUMAN	Q13495 homo sapien
20	39	36.8	739	1	SYG_HUMAN	P41250 homo sapien
21	38.5	36.3	246	1	TRY1_RAT	P00762 rattus norv
22	38.5	36.3	804	1	VP5_WTV	P12366 wound tumor
23	38	35.8	169	1	IL2_MOUSE	P04351 mus musculu
24	38	35.8	493	1	UDPH_YEAST	P38709 saccharomyc
25	38	35.8	517	1	FU26_YEAST	P31381 saccharomyc
26	38	35.8	553	1	MIS_RAT	P49000 rattus norv
27	38	35.8	555	1	MIS_MOUSE	P27106 mus musculu
28	38	35.8	622	1	MAK_MOUSE	Q04859 mus musculu
29	38	35.8	622	1	MAK_RAT	P20793 rattus norv
30	38	35.8	678	1	YIHQ_ECOLI	P32138 escherichia
31	38	35.8	857	1	AD22_MOUSE	Q9r1v6 mus musculu
32	38	35.8	996	1	ATA1_MAKNI	P70083 m sarcoplas
33	38	35.8	1041	1	EGT2_YEAST	P42835 saccharomyc
34	38	35.8	1069	1	ENTK_MOUSE	P97435 mus musculu
35	38	35.8	1075	1	FLO5_YEAST	P38894 saccharomyc
36	38	35.8	1116	1	MKH1_SCHPO	Q10407 schizosacch
37	38	35.8	1609	1	FIG2_YEAST	P25653 saccharomyc
38	38	35.8	1802	1	HKR1_YEAST	P41809 saccharomyc
39	38	35.8	2051	1	FAS1_YEAST	P07149 s fatty aci
40	37.5	35.4	336	1	CAHC_ARATH	P27140 arabidopsis
41	37.5	35.4	904	1	VGLB_HSV23	P06763 herpes simp
42	37.5	35.4	904	1	VGLB_HSV2H	P08666 herpes simp
43	37	34.9	100	1	MEN8_SILLA	O24356 silene lati
44	37	34.9	142	1	NIU2_RHOCA	Q10373 rhodobacter
45	37	34.9	151	1	YD36_HALN1	P20378 halobacteri

ALIGNMENTS

RESULT 1
FIL1_ANTMA
ID FIL1_ANTMA STANDARD; PRT; 99 AA.
AC Q38737;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Stamen-specific protein FIL1 precursor.
GN FIL1.
OS Antirrhinum majus (Garden snapdragon).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Lamiales; Veronicaceae; Antirrhinum.
OX NCBI_TaxID=4151;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Sippe 50;
RX MEDLINE=91375441; PubMed=1680216;
RA Nacken W.K.F., Huijser P., Beltran J.P., Saedler H., Sommer H.;
RT "Molecular characterization of two stamen-specific genes, tap1 and
RT fil1, that are expressed in the wild type, but not in the deficiens
RT mutant of Antirrhinum majus.";
RL Mol. Gen. Genet. 229:129-136(1991).
CC -!- TISSUE SPECIFICITY: STAMEN-SPECIFIC.
CC -!- SIMILARITY: BELONGS TO THE A9 / FIL1 FAMILY.
CC ---

CC ---
DR EMBL; X57296; CAA40553.1; -.
DR InterPro; IPR003612; AAI.
DR InterPro; IPR001768; Try/amyl_inhbtr.
DR Pfam; PF00234; tryp_alpha_amyl; 1.
DR SMART; SM00499; AAI; 1.
KW Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 99 STAMEN-SPECIFIC PROTEIN FIL1.
FT DISULFID 31 68 BY SIMILARITY.
FT DISULFID 41 57 BY SIMILARITY.
FT DISULFID 58 83 BY SIMILARITY.
FT DISULFID 70 90 BY SIMILARITY.
SQ SEQUENCE 99 AA; 10255 MW; 29A88517915BC0D6 CRC64;

Query Match 49.1%; Score 52; DB 1; Length 99;
Best Local Similarity 45.5%; Pred. No. 0.13;
Matches 10; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

```
Qy      1 SLVRLSSCVPVALMSAMTTSSS 22
          ||  |::| |  :: | || ||
Db     34 SLANLNACAPFVVLGAATTPSS 55
```

```
RESULT 2
GSN2_MOUSE
ID   GSN2_MOUSE      STANDARD;      PRT;   308 AA.
AC   Q9CY27;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Synaptic glycoprotein SC2.
GN   GPSN2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Embryonic liver;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RA   Strausberg R.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: BELONGS TO THE STEROID 5-ALPHA REDUCTASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
     or send an email to license@isb-sib.ch).
     ---------------------------------------------------------------------
     EMBL; AK010984; BAB27305.1; -.
DR   EMBL; BC019984; AAH19984.1; -.
DR   MGD; MGI:1915408; Gpsn2.
DR   InterPro; IPR001104; Strd5A_dhC.
DR   Pfam; PF02544; Steroid_dh; 1.
DR   PROSITE; PS50244; S5A_REDUCTASE; 1.
KW   Transmembrane; Glycoprotein.
FT   TRANSMEM      87    107       POTENTIAL.
FT   TRANSMEM     194    214       POTENTIAL.
FT   TRANSMEM     255    275       POTENTIAL.
FT   CARBOHYD     164    164       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     247    247       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   308 AA;  36090 MW;  0576C2813F2C5E4A CRC64;

  Query Match             38.7%;  Score 41;  DB 1;  Length 308;
  Best Local Similarity   72.7%;  Pred. No. 24;
  Matches    8;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       5 LSSCVPVALMS 15
            |: |||||| |
Db     261 LTQCVPVALFS 271


RESULT 3
UDPG_PYRPY
ID   UDPG_PYRPY      STANDARD;      PRT;   471 AA.
AC   O64459;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   UTP--glucose-1-phosphate uridylyltransferase (EC 2.7.7.9) (UDP-glucose
DE   pyrophosphorylase) (UDPGP) (UGPase).
OS   Pyrus pyrifolia (Japanese pear) (Pyrus serotina).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Rosales; Rosaceae; Maloideae; Pyrus.
OX   NCBI_TaxID=3767;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nijisseiki; TISSUE=Pollen;
RA   Kiyozumi D., Ishimizu T., Nakanishi T., Sakiyama F., Norioka S.;
RT   "Molecular cloning and nucleotide sequencing of a cDNA encoding UDP-
RT   glucose pyrophosphorylase of Japanese pear (Pyrus pyrifolia Nakai).";
RL   (In) Plant Gene Register PGR99-006.
CC   -!- FUNCTION: PLAYS A CENTRAL ROLE AS A GLUCOSYL DONOR IN CELLULAR
CC       METABOLIC PATHWAYS.
CC   -!- CATALYTIC ACTIVITY: UTP + alpha-D-glucose 1-phosphate =
CC       diphosphate + UDP-glucose.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE EUKARYOTIC UDPGP FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB013353; BAA25917.1; -.
DR   InterPro; IPR002618; UDPGP.
DR   Pfam; PF01704; UDPGP; 1.
KW   Transferase; Kinase; Nucleotidyltransferase.
SQ   SEQUENCE   471 AA;  51845 MW;  CE5523CE35E13B40 CRC64;

  Query Match             38.7%;  Score 41;  DB 1;  Length 471;
  Best Local Similarity   43.8%;  Pred. No. 36;
  Matches    7;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       4 RLSSCVPVALMSAMTT 19
            :  ||||: ||::  |
Db     126 KYGSCVPLLLMNSFNT 141


RESULT 4
SG2_RANRI
ID   SG2_RANRI       STANDARD;      PRT;   601 AA.
AC   P30945;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Secretogranin II precursor (SGII).
OS   Rana ridibunda (Laughing frog) (Marsh frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Neobatrachia; Ranoidea; Ranidae; Rana.
OX   NCBI_TaxID=8406;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pituitary;
RX   MEDLINE=96427274; PubMed=8830661;
RA   Anouar Y., Jegou S., Alexandre D., Lihrmann I., Conlon J.M.,
RA   Vaudry H.;
```

```
RT   "Molecular cloning of frog secretogranin II reveals the occurrence of
RT   several highly conserved potential regulatory peptides.";
RL   FEBS Lett. 394:295-299(1996).
RN   [2]
RP   SEQUENCE OF 183-215.
RC   TISSUE=Brain;
RX   MEDLINE=91285100; PubMed=2060624;
RA   Vaudry H., Conlon J.M.;
RT   "Identification of a peptide arising from the specific post-
RT   translation processing of secretogranin II.";
RL   FEBS Lett. 284:31-33(1991).
CC   -!- FUNCTION: MAY BE IMPORTANT IN REGULATION OF NEUROSECRETION.
CC   -!- SIMILARITY: BELONGS TO THE CHROMOGRANIN / SECRETOGRANIN PROTEIN
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U68757; AAB17470.1; -.
DR   PIR; S15867; S15867.
DR   InterPro; IPR001990; Granin.
DR   Pfam; PF01271; Granin; 1.
DR   PROSITE; PS00422; GRANINS_1; FALSE_NEG.
KW   Sulfation; Cleavage on pair of basic residues; Signal.
FT   SIGNAL        1     30       BY SIMILARITY.
FT   CHAIN        31    601       SECRETOGRANIN II.
FT   PEPTIDE     183    215       BRAIN PEPTIDE.
FT   MOD_RES     151    151       SULFATION (BY SIMILARITY).
SQ   SEQUENCE   601 AA;  69900 MW;  8D16FDAA1280A712 CRC64;

  Query Match             38.7%;  Score 41;  DB 1;  Length 601;
  Best Local Similarity   52.6%;  Pred. No. 46;
  Matches   10;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       5 LSSCVPVALMSAMTTSSSQ 23
           ||||: | |||    :| |
Db      13 LSSCILVILMSFSDAASFQ 31


RESULT 5
IL2_MIRAN
ID   IL2_MIRAN      STANDARD;      PRT;   154 AA.
AC   O62641;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Interleukin-2 precursor (IL-2) (T-cell growth factor) (TCGF).
GN   IL2.
OS   Mirounga angustirostris (Northern elephant seal).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Pinnipedia; Phocidae; Mirounga.
OX   NCBI_TaxID=9716;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98136706; PubMed=9476229;
RA   Shoda L.K.M., Brown W.C., Rice-Ficht A.C.;
RT   "Sequence and characterization of phocine interleukin 2.";
RL   J. Wildl. Dis. 34:81-90(1998).
CC   -!- FUNCTION: PRODUCED BY T-CELLS IN RESPONSE TO ANTIGENIC OR
CC       MITOGENIC STIMULATION, THIS PROTEIN IS REQUIRED FOR T-CELL
CC       PROLIFERATION AND OTHER ACTIVITIES CRUCIAL TO REGULATION OF THE
CC       IMMUNE RESPONSE. CAN STIMULATE B CELLS, MONOCYTES, LYMPHOKINE-
CC       ACTIVATED KILLER CELLS, NATURAL KILLER CELLS, AND GLIOMA CELLS (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: BELONGS TO THE IL-2 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U79187; AAC12258.1; -.
DR   HSSP; P01585; 3INK.
DR   InterPro; IPR000779; Interleukin-2.
DR   Pfam; PF00715; IL2; 1.
DR   PRINTS; PR00265; INTERLEUKIN2.
DR   ProDom; PD003649; Interleukin-2; 1.
DR   SMART; SM00189; IL2; 1.
DR   PROSITE; PS00424; INTERLEUKIN_2; 1.
KW   Cytokine; Glycoprotein; Immune response; Signal; Growth factor;
KW   T-cell.
FT   SIGNAL        1     20       BY SIMILARITY.
FT   CHAIN        21    154       INTERLEUKIN-2.
FT   CARBOHYD     23     23       O-LINKED (GALNAC. . .) (BY SIMILARITY).
FT   DISULFID     78    126       BY SIMILARITY.
SQ   SEQUENCE   154 AA;  17661 MW;  0C92337A4B16B6BB CRC64;

  Query Match             37.7%;  Score 40;  DB 1;  Length 154;
  Best Local Similarity   40.0%;  Pred. No. 17;
  Matches   10;  Conservative    8;  Mismatches    3;  Indels    4;  Gaps    1;

Qy       3 VRLSSCVPVALM----SAMTTSSSQ 23
           ::| ||: ::|:    || ||||::
Db       4 MQLLSCIALSLVLVANSAPTTSSTK 28


RESULT 6
Y412_MYCGE
ID   Y412_MYCGE     STANDARD;      PRT;   377 AA.
AC   P47652; Q49510;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical lipoprotein MG412 precursor.
GN   MG412.
OS   Mycoplasma genitalium.
OC   Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2097;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 33530 / G-37;
RX   MEDLINE=96026346; PubMed=7569993;
RA   Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA   Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA   Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA   Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA   Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA   Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT   "The minimal gene complement of Mycoplasma genitalium.";
RL   Science 270:397-403(1995).
RN   [2]
RP   SEQUENCE OF 1-74 AND 189-225 FROM N.A.
RC   STRAIN=ATCC 33530 / G-37;
RX   MEDLINE=94075230; PubMed=8253680;
RA   Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT   "A survey of the Mycoplasma genitalium genome by using random
RT   sequencing.";
RL   J. Bacteriol. 175:7918-7930(1993).
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC       (Potential).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; U39723; AAC71640.1; -.
DR   EMBL; U01702; AAB01014.1; ALT_INIT.
DR   EMBL; U02101; AAD12373.1; -.
DR   TIGR; MG412; -.
DR   InterPro; IPR002370; PstS.
DR   Pfam; PF01449; PstS; 1.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW   Hypothetical protein; Lipoprotein; Membrane; Signal;
KW   Complete proteome.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24    377       HYPOTHETICAL LIPOPROTEIN MG412.
FT   LIPID        24     24       N-ACYL DIGLYCERIDE (POTENTIAL).
SQ   SEQUENCE   377 AA;  42470 MW;  DD7E9BE19EBA70F2 CRC64;

  Query Match             37.7%;  Score 40;  DB 1;  Length 377;
  Best Local Similarity   47.4%;  Pred. No. 42;
  Matches    9;  Conservative    4;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       5 LSSCVPVALMSAMTTSSSQ 23
           || |  : |:||: :|| |
Db      21 LSGCANINLISAVGSSSVQ 39


RESULT 7
CAT2_CLOKL
ID   CAT2_CLOKL     STANDARD;      PRT;   429 AA.
AC   P38942;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   4-hydroxybutyrate coenzyme A transferase.
GN   CAT2.
OS   Clostridium kluyveri.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1534;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 555;
RA   Soehling B., Gottschalk G.;
RL   Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 149-429 FROM N.A.
RC   STRAIN=DSM 555;
RX   MEDLINE=96146540; PubMed=8550525;
RA   Soehling B., Gottschalk G.;
RT   "Molecular analysis of the anaerobic succinate degradation pathway in
RT   Clostridium kluyveri.";
RL   J. Bacteriol. 178:871-880(1996).
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; L21902; AAA92344.1; -.
DR   InterPro; IPR003702; AcetylCoA_hydro.
DR   Pfam; PF02550; AcetylCoA_hydro; 1.
KW   Transferase.
SQ   SEQUENCE   429 AA;  47071 MW;  71494C3666A8B52D CRC64;

  Query Match             37.7%;  Score 40;  DB 1;  Length 429;
  Best Local Similarity   38.1%;  Pred. No. 47;
  Matches    8;  Conservative    6;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 SLVRLSSCVPVALMSAMTTSS 21
           ::| ::||| | ||  : : |
Db     305 NMVSINSCVQVDLMGQVCSES 325


RESULT 8
CP1B_HUMAN
ID   CP1B_HUMAN     STANDARD;      PRT;   543 AA.
AC   Q16678; Q93089;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Cytochrome P450 1B1 (EC 1.14.14.1) (CYPIB1).
GN   CYP1B1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94230403; PubMed=8175734;
RA   Sutter T.R., Tang Y.M., Hayes C.L., Wo Y.-Y.P., Jabs E.W., Li X.,
RA   Yin H., Cody C.W., Greenlee W.F.;
RT   "Complete cDNA sequence of a human dioxin-inducible mRNA identifies a
RT   new gene subfamily of cytochrome P450 that maps to chromosome 2.";
RL   J. Biol. Chem. 269:13092-13099(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97067052; PubMed=8910454;
RA   Tang Y.M., Wo Y.-Y.P., Stewart J., Hawkins A.L., Griffin C.A.,
RA   Sutter T.R., Greenlee W.F.;
RT   "Isolation and characterization of the human cytochrome P450 CYP1B1
RT   gene.";
RL   J. Biol. Chem. 271:28324-28330(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RA   Strausberg R.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   VARIANTS GLC3A GLU-61; ASN-374 AND TRP-469.
RX   MEDLINE=98130535; PubMed=9463332;
RA   Bejjani B.A., Lewis R.A., Tomey K.F., Anderson K.L., Dueker D.K.,
RA   Jabak M., Astle W.F., Otterud B., Leppert M., Lupski J.R.;
RT   "Mutations in CYP1B1, the gene for cytochrome P4501B1, are the
RT   predominant cause of primary congenital glaucoma in Saudi Arabia.";
RL   Am. J. Hum. Genet. 62:325-333(1998).
RN   [5]
RP   VARIANT GLC3A TRP-365, AND VARIANTS.
RX   MEDLINE=98163441; PubMed=9497261;
RA   Stoilov I., Akarsu A.N., Alozie I., Child A., Barsoum-Homsy M.,
RA   Turacli M.E., Or M., Lewis R.A., Ozdemir N., Brice G., Aktan S.G.,
RA   Chevrette L., Coca-Prados M., Sarfarazi M.;
RT   "Sequence analysis and homology modeling suggest that primary
RT   congenital glaucoma on 2p21 results from mutations disrupting either
RT   the hinge region or the conserved core structures of cytochrome
RT   P4501B1.";
RL   Am. J. Hum. Genet. 62:573-584(1998).
RN   [6]
RP   VARIANTS LEU-432 AND SER-453.
RX   MEDLINE=99040639; PubMed=9823305;
RA   Bailey L.R., Roodi N., Dupont W.D., Parl F.F.;
RT   "Association of cytochrome P450 1B1 (CYP1B1) polymorphism with steroid
RT   receptor status in breast cancer.";
RL   Cancer Res. 58:5038-5041(1998).
RN   [7]
RP   ERRATUM.
RA   Bailey L.R., Roodi N., Dupont W.D., Parl F.F.;
RL   Cancer Res. 59:1388-1388(1999).
CC   -!- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
CC       MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN
CC       NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY
CC       OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY
CC       ACIDS, AND XENOBIOTICS.
CC   -!- FUNCTION: PARTICIPATES IN THE METABOLISM OF AN AS-YET-UNKNOWN
CC       BIOLOGICALLY ACTIVE MOLECULE THAT IS A PARTICIPANT IN EYE
```

```
CC       DEVELOPMENT.
CC    -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC    -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC    -!- TISSUE SPECIFICITY: EXPRESSED IN MANY TISSUES.
CC    -!- INDUCTION: BY POLYCYCLIC AROMATIC HYDROCARBONS (PAH) AND 2,3,7,8-
CC       TETRACHLORODIBENZO-P-DIOXIN (TCDD)..
CC    -!- DISEASE: DEFECTS IN CYP1B1 ARE A CAUSE OF PRIMARY CONGENITAL
CC       GLAUCOMA (GLC3A ALSO KNOW AS BUPHTHALMOS). THIS RECESSIVE DISEASE
CC       IS CHARACTERIZED BY LARGE OCULAR GLOBES RESULTING FROM INCREASED
CC       INTRAOCULAR PRESSURE. THE TERM BUPTHALMOS MEANS "OX EYE".
CC    -!- DISEASE: Defects in CYP1B1 are a cause of Peters anomaly, a
CC       congenital defect of the anterior chamber of the eye.
CC    -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC    -!- DATABASE: NAME=Cytochrome P450 Allele Nomenclature Committee;
CC       NOTE=CYP1B1 alleles;
CC       WWW="http://www.imm.ki.se/CYPalleles/cyp1b1.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U03688; AAA19567.1; -.
DR   EMBL; U56438; AAC50809.1; -.
DR   EMBL; BC012049; AAH12049.1; -.
DR   HSSP; P00179; 1DT6.
DR   Genew; HGNC:2597; CYP1B1.
DR   MIM; 601771; -.
DR   MIM; 231300; -.
DR   MIM; 604229; -.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   Pfam; PF00067; p450; 1.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW   Microsome; Endoplasmic reticulum; Disease mutation; Polymorphism.
FT   BINDING     470    470       HEME (BY SIMILARITY).
FT   VARIANT      48     48       R -> G (IN DBSNP:10012).
FT                                /FTId=VAR_011752.
FT   VARIANT      57     57       W -> C (IN CYP1B1*11).
FT                                /FTId=VAR_008350.
FT   VARIANT      61     61       G -> E (IN GLC3A; CYP1B1*12).
FT                                /FTId=VAR_001244.
FT   VARIANT     119    119       A -> S (IN DBSNP:1056827).
FT                                /FTId=VAR_011753.
FT   VARIANT     365    365       G -> W (IN GLC3A; CYP1B1*18).
FT                                /FTId=VAR_001245.
FT   VARIANT     374    374       D -> N (IN GLC3A).
FT                                /FTId=VAR_001246.
FT   VARIANT     379    379       P -> L (IN CYP1B1*19).
FT                                /FTId=VAR_008351.
FT   VARIANT     387    387       E -> K (IN CYP1B1*20).
FT                                /FTId=VAR_008352.
FT   VARIANT     390    390       R -> H (IN CYP1B1*21).
FT                                /FTId=VAR_008353.
FT   VARIANT     432    432       V -> L (IN CYP1B1*3).
FT                                /FTId=VAR_001248.
FT   VARIANT     437    437       P -> L (IN CYP1B1*23).
FT                                /FTId=VAR_008354.
FT   VARIANT     453    453       N -> S (IN CYP1B1*4).
FT                                /FTId=VAR_008355.
FT   VARIANT     469    469       R -> W (IN GLC3A; CYP1B1*25).
FT                                /FTId=VAR_001247.
SQ   SEQUENCE   543 AA;  60831 MW;  71BB6CA8055122DF CRC64;

  Query Match           37.7%;  Score 40;  DB 1;  Length 543;
  Best Local Similarity 45.0%;  Pred. No. 60;
  Matches    9;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       3 VRLSSCVPVALMSAMTTSSS 22
                 :| || ||| :  | | ::|
Db     389 MRFSSFVPVTIPHATTANTS 408


RESULT 9
AMYH_YEAST
ID   AMYH_YEAST     STANDARD;      PRT;  1367 AA.
AC   P08640; P08068;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Glucoamylase S1/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-
DE   glucosidase) (1,4-alpha-D-glucan glucohydrolase).
GN   STA1 OR STA2 OR MAL5 OR YIR019C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RA   Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA   Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA   Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA   Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA   Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA   Walsh S.V., Whitehead S.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX   MEDLINE=87194600; PubMed=3106330;
RA   Yamashita I., Nakamura M., Fukui S.;
RT   "Gene fusion is a possible mechanism underlying the evolution of
RT   STA1.";
RL   J. Bacteriol. 169:2142-2149(1987).
RN   [3]
RP   SEQUENCE OF 1-31 FROM N.A.
RC   STRAIN=SPX101-1C;
RX   MEDLINE=89031230; PubMed=3141213;
RA   Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT   "Similar short elements in the 5' regions of the STA2 and SGA genes
RT   from Saccharomyces cerevisiae.";
RL   FEBS Lett. 239:179-184(1988).
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC       glucose residues successively from non-reducing ends of the chains
CC       with release of beta-D-glucose.
CC   -!- SIMILARITY: TO S.POMBE SPBC215.13.
CC   -!- SIMILARITY: SOME, TO S.POMBE SPCC285.13C.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z38061; CAA86176.1; -.
DR   EMBL; M16164; AAA35014.1; -.
DR   EMBL; M16165; AAA35015.1; -.
DR   EMBL; X13857; CAA32069.1; -.
DR   PIR; B26877; B26877.
DR   PIR; A26877; A26877.
DR   PIR; S48478; S48478.
DR   SGD; S0001458; MUC1.
KW   Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW   Signal; Multigene family.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   1367       GLUCOAMYLASE S1/S2.
FT   DOMAIN      210   1367       SER/THR-RICH.
FT   CARBOHYD    817    817       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    874    874       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1367 AA;  136110 MW;  91C00E2DBD61AA9D CRC64;
```

```
  Query Match             37.7%;  Score 40;  DB 1;  Length 1367;
  Best Local Similarity   58.8%;  Pred. No. 1.5e+02;
  Matches   10;  Conservative    1;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       6 SSCVPVALMSAMTTSSS 22
           || |||   |: || ||
Db     834 SSSVPVPTPSSSTTESS 850


RESULT 10
ALA1_CANAL
ID   ALA1_CANAL     STANDARD;      PRT;  1419 AA.
AC   O13368;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Agglutinin-like protein ALA1 precursor (Agglutinin-like adhesin).
GN   ALA1 OR ALS5.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98053977; PubMed=9393828;
RA   Gaur N.K., Klotz S.A.;
RT   "Expression, cloning, and characterization of a Candida albicans
RT   gene, ALA1, that confers adherence properties upon Saccharomyces
RT   cerevisiae for extracellular matrix proteins.";
RL   Infect. Immun. 65:5289-5294(1997).
CC   -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
CC   -!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF025429; AAB88883.1; -.
KW   Cell adhesion; Glycoprotein; Repeat; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18   1419       AGGLUTININ-LIKE PROTEIN ALA1.
FT   DOMAIN      399    404       POLY-THR.
FT   DOMAIN      408    418       POLY-THR.
FT   DOMAIN      437    441       POLY-THR.
FT   DOMAIN      673    676       POLY-SER.
FT   DOMAIN      687    690       POLY-SER.
FT   DOMAIN      700    703       POLY-SER.
     DOMAIN      719    724       POLY-SER.
     DOMAIN      749    752       POLY-SER.
     DOMAIN      787    791       POLY-SER.
FT   DOMAIN      869    872       POLY-SER.
FT   DOMAIN      875    883       POLY-SER.
FT   DOMAIN      901    911       POLY-SER.
FT   DOMAIN     1216   1221       POLY-SER.
FT   CARBOHYD    665    665       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    919    919       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1301   1301       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1326   1326       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1419 AA;  149635 MW;  249F33F688A9D5B6 CRC64;

  Query Match             37.7%;  Score 40;  DB 1;  Length 1419;
  Best Local Similarity   50.0%;  Pred. No. 1.6e+02;
  Matches   11;  Conservative    3;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       1 SLVRLSSCVPVALMSAMTTSSS 22
           |:| |||   : | | | :|||
Db     682 SIVGLSSSSDIPLSSDMPSSSS 703


RESULT 11
L302_BOMMO
ID   L302_BOMMO     STANDARD;      PRT;   263 AA.
AC   Q00801;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Low molecular mass 30 kDa lipoprotein 21G1 precursor.
GN   21G1.
OS   Bombyx mori (Silk moth).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
OC   Ditrysia; Bombycoidea; Bombycidae; Bombyx.
OX   NCBI_TaxID=7091;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Tokai X Asahi; TISSUE=Fat body;
RX   MEDLINE=91355222; PubMed=1883839;
RA   Mori S., Izumi S., Tomino S.;
RT   "Complete nucleotide sequences of major plasma protein genes of
RT   Bombyx mori.";
RL   Biochim. Biophys. Acta 1090:129-132(1991).
CC   -!- FUNCTION: NOT KNOWN. THIS LIPOPROTEIN BELONGS TO THE GROUP OF
CC       STRUCTURALLY RELATED '30 KDA PROTEINS' THAT COMPRISE MAJOR PROTEIN
CC       COMPONENTS OF THE FIFTH (AND LAST) INSTAR LARVAE AND OF PUPAE.
CC   -!- TISSUE SPECIFICITY: LARVAL HEMOLYMPH; OOCYTE.
CC   -!- INDUCTION: SURGICAL EXTIRPATION OF THE CORPORA ALLATA.
CC   -!- SIMILARITY: TO OTHER 30 kDa PROTEINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X54734; CAA38531.1; -.
DR   PIR; S17661; S17661.
DR   InterPro; IPR004943; Lipoprotein_11.
DR   Pfam; PF03260; Lipoprotein_11; 1.
KW   Hemolymph; Lipoprotein; Signal.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24    263       LOW MOLECULAR MASS 30 KDA LIPOPROTEIN
FT                                21G1.
SQ   SEQUENCE   263 AA;  30183 MW;  6F62AFD48BE6CF5B CRC64;

  Query Match             37.3%;  Score 39.5;  DB 1;  Length 263;
  Best Local Similarity   50.0%;  Pred. No. 35;
  Matches   13;  Conservative    3;  Mismatches    5;  Indels    5;  Gaps    1;

Qy       2 LVRLSSCV-----PVALMSAMTTSSS 22
           ||  :|||      || |||:: |||
Db       4 LVVFASCVLAVSAGVAEMSAVSMSSS 29


RESULT 12
RRPO_LYCVA
ID   RRPO_LYCVA     STANDARD;      PRT;  2210 AA.
AC   P14240;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   RNA polymerase (EC 2.7.7.48).
GN   L.
OS   Lymphocytic choriomeningitis virus (strain Armstrong).
OC   Viruses; ssRNA negative-strand viruses; Arenaviridae; Arenavirus.
OX   NCBI_TaxID=11624;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89204909; PubMed=2705303;
RA   Salvato M.S., Shimomaye E.M., Oldstone M.B.A.;
```

```
RT   "The primary structure of the lymphocytic choriomeningitis virus L
RT   gene encodes a putative RNA polymerase.";
RL   Virology 169:377-384(1989).
RN   [2]
RP   SEQUENCE OF 161-387; 424-619 AND 1646-1906 FROM N.A.
RX   MEDLINE=88072084; PubMed=3318094;
RA   Singh M.K., Fuller-Pace F.V., Buchmeier M.J., Southern P.J.;
RT   "Analysis of the genomic L RNA segment from lymphocytic
RT   choriomeningitis virus.";
RL   Virology 161:448-456(1987).
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       (RNA)(N).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04331; AAA66591.1; -.
DR   EMBL; M18381; AAA46258.1; ALT_SEQ.
DR   EMBL; M18382; AAA46259.1; -.
DR   EMBL; M18383; AAA46260.1; ALT_SEQ.
DR   PIR; A30181; RRXPLC.
KW   RNA-directed RNA polymerase; Transferase.
FT   CONFLICT    164    164       L -> Y (IN REF. 2).
FT   CONFLICT    354    354       Q -> R (IN REF. 2).
FT   CONFLICT    361    361       K -> E (IN REF. 2).
FT   CONFLICT    382    382       H -> D (IN REF. 2).
FT   CONFLICT    552    552       C -> S (IN REF. 2).
FT   CONFLICT   1727   1727       R -> L (IN REF. 2).
SQ   SEQUENCE   2210 AA;  254529 MW;  470C8E623176AFD3 CRC64;

  Query Match             37.3%;  Score 39.5;  DB 1;  Length 2210;
  Best Local Similarity   61.9%;  Pred. No. 2.9e+02;
  Matches   13;  Conservative    3;  Mismatches    4;  Indels    1;  Gaps    1;

Qy       1 SLVRLSSCVPVALMSAMTTSS 21
           || |||| | :|| ::| |||
Db     486 SLRRLSS-VCLALTNSMKTSS 505


RESULT 13
GSN2_RAT
ID   GSN2_RAT       STANDARD;      PRT;   308 AA.
AC   Q64232;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Synaptic glycoprotein SC2.
GN   GPSN2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=93021239; PubMed=1404491;
RA   Johnston I.G., Rush S.J., Gurd J.W., Brown I.R.;
RT   "Molecular cloning of a novel mRNA using an antibody directed against
RT   synaptic glycoproteins.";
RL   J. Neurosci. Res. 32:159-166(1992).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- TISSUE SPECIFICITY: Expressed at high levels in brain and is also
CC       found at lower levels in several other tissues.
CC   -!- PTM: Glycosylated.
CC   -!- SIMILARITY: BELONGS TO THE STEROID 5-ALPHA REDUCTASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; S45663; AAB23534.1; -.
DR   InterPro; IPR001104; Strd5A_dhC.
DR   Pfam; PF02544; Steroid_dh; 1.
DR   PROSITE; PS50244; S5A_REDUCTASE; 1.
KW   Transmembrane; Glycoprotein.
FT   TRANSMEM     87    107       POTENTIAL.
FT   TRANSMEM    194    214       POTENTIAL.
FT   TRANSMEM    255    275       POTENTIAL.
FT   CARBOHYD    164    164       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    247    247       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   308 AA;  36122 MW;  9E1B6280F61DD463 CRC64;

  Query Match             36.8%;  Score 39;  DB 1;  Length 308;
  Best Local Similarity   63.6%;  Pred. No. 49;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       5 LSSCVPVALMS 15
           :: |||||| |
Db     261 MTQCVPVALFS 271


RESULT 14
CAR2_RHINI
ID   CAR2_RHINI     STANDARD;      PRT;   391 AA.
AC   P43231;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Rhizopuspepsin 2 precursor (EC 3.4.23.21) (Aspartate protease).
OS   Rhizopus niveus.
OC   Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
OC   Rhizopus.
OX   NCBI_TaxID=4844;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Yamazaki / IFO 4810;
RA   Horiuchi H., Nakamura H., Okazaki T., Yano K., Takagi M.;
RL   Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity
CC       similar to that of pepsin A, preferring hydrophobic residues at P1
CC       and P1'. Clots milk and activates trypsinogen. Does not cleave 4-
CC       Gln-|-His-5, but does cleave 10-His-|-Leu-11 and 12-Val-|-Glu-13
CC       in B chain of insulin.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X56964; CAA40284.1; -.
DR   HSSP; P06026; 2APR.
DR   MEROPS; A01.012; -.
DR   InterPro; IPR001461; Aspproteasea1.
DR   InterPro; IPR001969; Aspprotease_site.
DR   Pfam; PF00026; asp; 1.
DR   PRINTS; PR00792; PEPSIN.
DR   PROSITE; PS00141; ASP_PROTEASE; 2.
KW   Hydrolase; Aspartyl protease; Zymogen; Signal; Multigene family.
FT   SIGNAL       1     21       POTENTIAL.
FT   PROPEP      22     68       POTENTIAL.
FT   CHAIN       69    391       RHIZOPUSPEPSIN 2.
FT   ACT_SITE   102    102       BY SIMILARITY.
FT   ACT_SITE   285    285       BY SIMILARITY.
```

```
FT   DISULFID    115    118       BY SIMILARITY.
FT   DISULFID    319    352       BY SIMILARITY.
SQ   SEQUENCE   391 AA;  41234 MW;  3F34230659DEA960 CRC64;

  Query Match             36.8%;  Score 39;  DB 1;  Length 391;
  Best Local Similarity   40.0%;  Pred. No. 63;
  Matches    8;  Conservative    5;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       2 LVRLSSCVPVALMSAMTTSS 21
           |  :|||| :| |:  | ::
Db       3 LTLISSCVALAFMALATEAA 22


RESULT 15
MMM1_YEAST
ID   MMM1_YEAST     STANDARD;      PRT;   426 AA.
AC   P41800;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Mitochondrial outer membrane protein MMM1.
GN   MMM1 OR YLL006W OR L1357.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YH8;
RX   MEDLINE=94375516; PubMed=8089172;
RA   Burgess S.M., Delannoy M., Jensen R.E.;
RT   "MMM1 encodes a mitochondrial outer membrane protein essential for
RT   establishing and maintaining the structure of yeast mitochondria.";
RL   J. Cell Biol. 126:1375-1391(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / FY23;
RX   MEDLINE=96405918; PubMed=8810043;
RA   Miosga T., Zimmermann F.K.;
RT   "Sequence analysis of the CEN12 region of Saccharomyces cerevisiae on
RT   a 43.7 kb fragment of chromosome XII including an open reading frame
RT   homologous to the human cystic fibrosis transmembrane conductance
RT   regulator protein CFTR.";
RL   Yeast 12:693-708(1996).
CC   -!- FUNCTION: ESSENTIAL FOR ESTABLISHING AND MAINTAINING THE STRUCTURE
CC       OF MITOCHONDRIA. MAY MAINTAIN MITOCHONDRIA IN AN ELONGATED SHAPE
CC       BY ATTACHING THE MITOCHONDRION TO AN EXTERNAL FRAMEWORK, SUCH AS
CC       THE CYTOSKELETON.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC       outer membrane.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
     between the Swiss Institute of Bioinformatics and the EMBL outstation -
     the European Bioinformatics Institute. There are no restrictions on its
     use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L32793; AAA53581.1; -.
DR   EMBL; X91488; CAA62763.1; -.
DR   EMBL; Z73111; CAA97449.1; -.
DR   SGD; S0003929; MMM1.
KW   Mitochondrion; Outer membrane; Transmembrane.
FT   DOMAIN        1     91       INTERMEMBRANE (POTENTIAL).
FT   TRANSMEM     92    116       POTENTIAL.
FT   DOMAIN      117    426       CYTOPLASMIC (POTENTIAL).
FT   CONFLICT    200    200       W -> S (IN REF. 1).
FT   CONFLICT    225    225       S -> W (IN REF. 1).
FT   CONFLICT    357    357       S -> A (IN REF. 1).
FT   CONFLICT    365    365       E -> Q (IN REF. 1).
FT   CONFLICT    368    368       S -> P (IN REF. 1).
SQ   SEQUENCE   426 AA;  48660 MW;  3610AEC2109355AF CRC64;

  Query Match             36.8%;  Score 39;  DB 1;  Length 426;
  Best Local Similarity   36.4%;  Pred. No. 68;
  Matches    8;  Conservative    7;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 SLVRLSSCVPVALMSAMTTSSS 22
           |:||  :|: |:| :|   :|:
Db     313 SIVRFQACLTVSLTNAEEFAST 334


Search completed: May  7, 2003, 09:30:22
Job time : 14 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: May 7, 2003, 09:28:35 ; Search time 29 Seconds
(without alignments)
163.417 Million cell updates/sec

Title: US-09-674-973A-17
Perfect score: 106
Sequence: 1 SLVRLSSCVPVALMSAMTTSSSQ 23

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	47.5	44.8	200	12	O41194	O41194 porcine rep
2	47	44.3	348	8	Q94W44	Q94w44 gnatholepis
3	47	44.3	412	12	Q8QRX0	Q8qrx0 chimpanzee
4	46	43.4	196	8	Q9MM83	Q9mm83 darevskia p
5	46	43.4	198	8	Q9MDL2	Q9mdl2 darevskia m
6	46	43.4	198	8	O21629	O21629 darevskia c
7	46	43.4	198	8	O21624	O21624 darevskia c
8	46	43.4	198	8	O21625	O21625 darevskia d
9	46	43.4	198	8	O21627	O21627 darevskia m
10	46	43.4	198	8	Q9MNY4	Q9mny4 lacerta arm
11	46	43.4	198	8	Q9MNY3	Q9mny3 lacerta dah
12	46	43.4	198	8	Q9MM88	Q9mm88 darevskia d
13	46	43.4	198	8	Q9MM86	Q9mm86 darevskia r
14	46	43.4	198	8	Q9MM85	Q9mm85 darevskia b
15	46	43.4	230	16	Q8XHZ2	Q8xhz2 clostridium
16	45.5	42.9	286	16	O53852	O53852 mycobacteri
17	45	42.5	102	16	O50482	O50482 streptomyce
18	45	42.5	198	8	Q9MM84	Q9mm84 darevskia l
19	45	42.5	698	3	Q9P8D7	Q9p8d7 candida clo
20	45	42.5	1388	5	Q8WQ36	Q8wq36 leishmania
21	44	41.5	238	5	Q8SSM9	Q8ssm9 dictyosteli
22	43	40.6	166	5	Q8SU40	Q8su40 encephalito
23	43	40.6	198	8	Q9MM89	Q9mm89 darevskia s
24	43	40.6	227	16	P94413	P94413 bacillus su
25	43	40.6	231	8	O98992	O98992 heterosigma
26	43	40.6	251	17	Q9YAT4	Q9yat4 aeropyrum p
27	43	40.6	273	16	Q9CIN0	Q9cin0 lactococcus
28	43	40.6	284	16	Q8ZPT1	Q8zptl salmonella
29	43	40.6	293	2	Q8RLY2	Q8rly2 salmonella
30	43	40.6	314	10	Q9LM18	Q9lm18 arabidopsis
31	43	40.6	322	13	Q9PTU0	Q9ptu0 brachydanio
32	43	40.6	323	12	Q9J1B0	Q9j1b0 nelson bay
33	43	40.6	337	16	Q92TF5	Q92tf5 rhizobium m
34	43	40.6	970	16	Q9A7U6	Q9a7u6 caulobacter
35	42	39.6	269	5	Q9VVW9	Q9vvw9 drosophila
36	42	39.6	539	5	Q9U149	Q9u149 leishmania
37	42	39.6	625	5	Q9W2V9	Q9w2v9 drosophila
38	42	39.6	945	3	Q8X1V8	Q8x1v8 aspergillus
39	41	38.7	96	10	Q9S7A3	Q9s7a3 arabidopsis
40	41	38.7	134	5	Q8SXV5	Q8sxv5 drosophila
41	41	38.7	144	16	Q92LN5	Q92ln5 rhizobium m
42	41	38.7	149	12	Q81753	Q81753 hepatitis c
43	41	38.7	155	6	Q9XT83	Q9xt83 halichoerus
44	41	38.7	157	5	Q9VA41	Q9va41 drosophila
45	41	38.7	191	11	Q9D9Q8	Q9d9q8 mus musculu

ALIGNMENTS

RESULT 1
O41194
ID O41194 PRELIMINARY; PRT; 200 AA.
AC O41194;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Envelope protein.
GN ENV.
OS Porcine reproductive and respiratory syndrome virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Arteriviridae; Arterivirus.
OX NCBI_TaxID=28344;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NADC-9;
RX MEDLINE=97335197; PubMed=9191863;
RA Andreyev V.G., Wesley R.D., Mengeling W.L., Vorwald A.C., Lager K.M.;
RT "Genetic variation and phylogenetic relationships of 22 porcine
RT reproductive and respiratory syndrome virus (PRRSV) field strains
RT based on sequence analysis of open reading frame 5.";
RL Arch. Virol. 142:993-1001(1997).
DR EMBL; U66393; AAC57967.1; -.
DR InterPro; IPR001332; Arteri_glycop.
DR InterPro; IPR003239; Porcine_RR_virus.
DR Pfam; PF00951; Arteri_glycop; 1.
DR ProDom; PD001151; Porcine_RR_virus; 1.
SQ SEQUENCE 200 AA; 22248 MW; 076D0E27849377A6 CRC64;

Query Match 44.8%; Score 47.5; DB 12; Length 200;
Best Local Similarity 72.2%; Pred. No. 4.7;
Matches 13; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

```
Qy       5 LSSCVPVALMSAMTTSSS 22
           :|||  |||:|| |||||
Db      21 VSSCF-VALVSANTTSSS 37
```

RESULT 2

```
Q94W44
ID   Q94W44      PRELIMINARY;      PRT;   348 AA.
AC   Q94W44;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   NADH dehydrogenase subunit 2.
OS   Gnatholepis scapulostigma (shoulderspot goby).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Gobioidei;
OC   Gobiidae; Gnatholepis.
OX   NCBI_TaxID=166749;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GNATHSCAP;
RA   Thacker C.E.;
RT   "Molecular Phylogeny of the Gobioid Fishes.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
DR   EMBL; AF391520; AAL16621.1; -.
DR   InterPro; IPR001750; Oxidored_q1.
DR   Pfam; PF00361; oxidored_q1; 1.
KW   Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
SQ   SEQUENCE   348 AA;  37997 MW;  F96C513FDB4B73F3 CRC64;

  Query Match             44.3%;  Score 47;  DB 8;  Length 348;
  Best Local Similarity   57.1%;  Pred. No. 9.9;
  Matches   12;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 SLVRLSSCVPVALMSAMTTSS 21
           || ||||  |  |:|| | |:
Db     312 SLWRLSSARPTILLSASTLSA 332


RESULT 3
Q8QRX0
ID   Q8QRX0      PRELIMINARY;      PRT;   412 AA.
AC   Q8QRX0;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Glycoprotein UL139.
OS   chimpanzee cytomegalovirus.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Betaherpesvirinae; Cytomegalovirus.
OX   NCBI_TaxID=188763;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Davison A.J., Akter P., Dolan A., Wright K.M., Addison C.,
RA   Alcendor D.J., Hayward G.S., McGeoch D.J.;
     "The human cytomegalovirus genome revisited.";
     Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
     EMBL; AF480884; AAM00767.1; -.
SQ   SEQUENCE   412 AA;  44758 MW;  83A134FD8372CB76 CRC64;

  Query Match             44.3%;  Score 47;  DB 12;  Length 412;
  Best Local Similarity   54.5%;  Pred. No. 12;
  Matches   12;  Conservative    2;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       1 SLVRLSSCVPVALMSAMTTSSS 22
           :|| ||| |  || |  || :|
Db     260 TLVALSSAVSAALASETTTGTS 281


RESULT 4
Q9MM83
ID   Q9MM83      PRELIMINARY;      PRT;   196 AA.
AC   Q9MM83;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ATP synthase A chain (EC 3.6.1.34) (Fragment).
OS   Darevskia parvula.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Scincomorpha; Lacertoidea;
OC   Lacertidae; Darevskia.
OX   NCBI_TaxID=122336;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Fu J.;
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A
CC       DIRECT ROLE IN THE TRANSLOCATION OF PROTONS ACROSS THE MEMBRANE
CC       (BY SIMILARITY).
CC   -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC       CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC       SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC       HAS THREE MAIN SUBUNITS: A, B AND C (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
DR   EMBL; AF206170; AAF70416.1; -.
DR   InterPro; IPR000568; ATPsynt_Asub.
DR   Pfam; PF00119; ATP-synt_A; 1.
DR   PRINTS; PR00123; ATPASEA.
DR   TIGRFAMs; TIGR01131; ATP_synt_6_or_A; 1.
DR   PROSITE; PS00449; ATPASE_A; 1.
KW   CF(0); Hydrogen ion transport; Mitochondrion; Transmembrane.
FT   NON_TER       1       1
FT   NON_TER     196     196
SQ   SEQUENCE   196 AA;  21645 MW;  241F99DED86C0778 CRC64;

  Query Match             43.4%;  Score 46;  DB 8;  Length 196;
  Best Local Similarity   45.0%;  Pred. No. 8.2;
  Matches    9;  Conservative    7;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 LVRLSSCVPVALMSAMTTSS 21
           |::|:|   :|||: |||::
Db     159 LIQLTSTAVLALMNTMTTTA 178


RESULT 5
Q9MDL2
ID   Q9MDL2      PRELIMINARY;      PRT;   198 AA.
AC   Q9MDL2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ATP synthase A chain (EC 3.6.1.34) (Fragment).
OS   Darevskia mixta.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Scincomorpha; Lacertoidea;
OC   Lacertidae; Darevskia.
OX   NCBI_TaxID=122392;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Fu J., Murphy R.W., Darevsky I.S.;
RT   "Limited genetic variation in Lacerta mixta and its parthenogenetic
RT   daughter species: evidence from cytochrome b and ATPase 6 gene DNA
RT   sequences.";
RL   Genetica 0:0-0(1999).
CC   -!- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A
CC       DIRECT ROLE IN THE TRANSLOCATION OF PROTONS ACROSS THE MEMBRANE
CC       (BY SIMILARITY).
CC   -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC       CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC       SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC       HAS THREE MAIN SUBUNITS: A, B AND C (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
DR   EMBL; AF147803; AAF73119.1; -.
DR   EMBL; AF147802; AAF73118.1; -.
DR   InterPro; IPR000568; ATPsynt_Asub.
```

```
DR   Pfam; PF00119; ATP-synt_A; 1.
DR   PRINTS; PR00123; ATPASEA.
DR   TIGRFAMs; TIGR01131; ATP_synt_6_or_A; 1.
DR   PROSITE; PS00449; ATPASE_A; 1.
KW   CF(0); Hydrogen ion transport; Mitochondrion; Transmembrane.
FT   NON_TER       1      1
FT   NON_TER     198    198
SQ   SEQUENCE   198 AA;  21766 MW;  B60A9F0B32B07DCF CRC64;

  Query Match             43.4%;  Score 46;  DB 8;  Length 198;
  Best Local Similarity   45.0%;  Pred. No. 8.3;
  Matches    9;  Conservative    7;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 LVRLSSCVPVALMSAMTTSS 21
           |::|:|   :|||: |||::
Db     161 LIQLTSTAVLALMNTMTTTA 180


RESULT 6
O21629
ID   O21629      PRELIMINARY;      PRT;   198 AA.
AC   O21629; Q9MM87;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ATP synthase A chain (EC 3.6.1.34) (Fragment).
OS   Darevskia clarkorum.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Scincomorpha; Lacertoidea;
OC   Lacertidae; Darevskia.
OX   NCBI_TaxID=122333;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Fu J., Murphy R.W., Darevsky I.S.;
RT   "Towards the phylogeny of Caucasian rock lizards: implications from
RT   mitochondrial DNA gene sequences (Reptilia: Lacertidae).";
RL   Zool. J. Linn. Soc. 121:463-477(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Fu J., Murphy R.W.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Fu J., Murphy R.W.;
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Fu J.;
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A
CC       DIRECT ROLE IN THE TRANSLOCATION OF PROTONS ACROSS THE MEMBRANE
CC       (BY SIMILARITY).
CC   -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC       CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC       SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC       HAS THREE MAIN SUBUNITS: A, B AND C (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
DR   EMBL; U88598; AAB65099.2; -.
DR   EMBL; AF206166; AAF70412.1; -.
DR   InterPro; IPR000568; ATPsynt_Asub.
DR   Pfam; PF00119; ATP-synt_A; 1.
DR   PRINTS; PR00123; ATPASEA.
DR   TIGRFAMs; TIGR01131; ATP_synt_6_or_A; 1.
DR   PROSITE; PS00449; ATPASE_A; 1.
KW   CF(0); Hydrogen ion transport; Mitochondrion; Transmembrane.
FT   NON_TER       1      1
FT   NON_TER     198    198
SQ   SEQUENCE   198 AA;  21784 MW;  759C72A79C691087 CRC64;

  Query Match             43.4%;  Score 46;  DB 8;  Length 198;
  Best Local Similarity   45.0%;  Pred. No. 8.3;
  Matches    9;  Conservative    7;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 LVRLSSCVPVALMSAMTTSS 21
           |::|:|   :|||: |||::
Db     161 LIQLTSTAVLALMNTMTTTA 180


RESULT 7
O21624
ID   O21624      PRELIMINARY;      PRT;   198 AA.
AC   O21624;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ATP synthase A chain (EC 3.6.1.34) (Fragment).
OS   Darevskia caucasica.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Scincomorpha; Lacertoidea;
OC   Lacertidae; Darevskia.
OX   NCBI_TaxID=122391;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Fu J., Murphy R.W., Darevsky I.S.;
RT   "Towards the phylogeny of Caucasian rock lizards: implications from
RT   mitochondrial DNA gene sequences (Reptilia: Lacertidae).";
RL   Zool. J. Linn. Soc. 121:463-477(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Fu J., Murphy R.W.;
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A
CC       DIRECT ROLE IN THE TRANSLOCATION OF PROTONS ACROSS THE MEMBRANE
CC       (BY SIMILARITY).
CC   -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC       CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC       SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC       HAS THREE MAIN SUBUNITS: A, B AND C (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
DR   EMBL; U88593; AAB65098.2; -.
DR   InterPro; IPR000568; ATPsynt_Asub.
DR   Pfam; PF00119; ATP-synt_A; 1.
DR   PRINTS; PR00123; ATPASEA.
DR   TIGRFAMs; TIGR01131; ATP_synt_6_or_A; 1.
DR   PROSITE; PS00449; ATPASE_A; 1.
KW   CF(0); Hydrogen ion transport; Mitochondrion; Transmembrane.
FT   NON_TER       1      1
FT   NON_TER     198    198
SQ   SEQUENCE   198 AA;  21815 MW;  1DAEB1AAEDCED167 CRC64;

  Query Match             43.4%;  Score 46;  DB 8;  Length 198;
  Best Local Similarity   45.0%;  Pred. No. 8.3;
  Matches    9;  Conservative    7;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 LVRLSSCVPVALMSAMTTSS 21
           |::|:|   :|||: |||::
Db     161 LIQLTSTAVLALMNTMTTTA 180


RESULT 8
O21625
ID   O21625      PRELIMINARY;      PRT;   198 AA.
AC   O21625;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ATP synthase A chain (EC 3.6.1.34) (Fragment).
OS   Darevskia daghestanica.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Scincomorpha; Lacertoidea;
OC   Lacertidae; Darevskia.
```

```
OX   NCBI_TaxID=122350;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Fu J., Murphy R.W., Darevsky I.S.;
RT   "Towards the phylogeny of Caucasian rock lizards: implications from
RT   mitochondrial DNA gene sequences (Reptilia: Lacertidae).";
RL   Zool. J. Linn. Soc. 121:463-477(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Fu J., Murphy R.W.;
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A
CC       DIRECT ROLE IN THE TRANSLOCATION OF PROTONS ACROSS THE MEMBRANE
CC       (BY SIMILARITY).
CC   -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC       CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC       SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC       HAS THREE MAIN SUBUNITS: A, B AND C (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
DR   EMBL; U88594; AAB65103.2; -.
DR   HSSP; P00855; 1C17.
DR   InterPro; IPR000568; ATPsynt_Asub.
DR   Pfam; PF00119; ATP-synt_A; 1.
DR   PRINTS; PR00123; ATPASEA.
DR   TIGRFAMs; TIGR01131; ATP_synt_6_or_A; 1.
DR   PROSITE; PS00449; ATPASE_A; 1.
KW   CF(0); Hydrogen ion transport; Mitochondrion; Transmembrane.
FT   NON_TER       1      1
FT   NON_TER     198    198
SQ   SEQUENCE   198 AA;  21778 MW;  40EF8612D318AFCF CRC64;

  Query Match             43.4%;  Score 46;  DB 8;  Length 198;
  Best Local Similarity   45.0%;  Pred. No. 8.3;
  Matches    9;  Conservative    7;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 LVRLSSCVPVALMSAMTTSS 21
           |::|:|   :|||: |||::
Db     161 LIQLTSTAVLALMNTMTTTA 180


RESULT 9
O21627
ID   O21627      PRELIMINARY;      PRT;   198 AA.
AC   O21627;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ATPase 6 (Fragment).
OS   Darevskia mixta.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
     Lepidosauria; Squamata; Scleroglossa; Scincomorpha; Lacertoidea;
     Lacertidae; Darevskia.
     NCBI_TaxID=122392;
RN   [1]
RP   SEQUENCE OF 16-129 FROM N.A.
RA   Fu J., Murphy R.W., Darevsky I.S.;
RT   "Towards the phylogeny of Caucasian rock lizards: implications from
RT   mitochondrial DNA gene sequences (Reptilia: Lacertidae).";
RL   Zool. J. Linn. Soc. 121:463-477(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Fu J., Murphy R.W., Darevsky I.S.;
RT   "Limited genetic variation in Lacerta mixta and its parthenogenetic
RT   daughter species: evidence from cytochrome b and ATPase 6 gene DNA
RT   sequences.";
RL   Genetica 0:0-0(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Fu J.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF147801; AAF73117.1; -.
DR   InterPro; IPR000568; ATPsynt_Asub.
DR   Pfam; PF00119; ATP-synt_A; 1.
DR   TIGRFAMs; TIGR01131; ATP_synt_6_or_A; 1.
DR   PROSITE; PS00449; ATPASE_A; 1.
KW   Mitochondrion.
FT   NON_TER       1      1
FT   NON_TER     198    198
SQ   SEQUENCE   198 AA;  21752 MW;  B6181D9B2032EDCF CRC64;

  Query Match             43.4%;  Score 46;  DB 8;  Length 198;
  Best Local Similarity   45.0%;  Pred. No. 8.3;
  Matches    9;  Conservative    7;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 LVRLSSCVPVALMSAMTTSS 21
           |::|:|   :|||: |||::
Db     161 LIQLTSTAVLALMNTMTTTA 180


RESULT 10
Q9MNY4
ID   Q9MNY4      PRELIMINARY;      PRT;   198 AA.
AC   Q9MNY4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ATP synthase A chain (EC 3.6.1.34) (Fragment).
OS   Lacerta armeniaca.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Scincomorpha; Lacertoidea;
OC   Lacertidae; Lacerta.
OX   NCBI_TaxID=94909;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Fu J., Murphy R.W., Darevsky I.S.;
RT   "Limited genetic variation in Lacerta mixta and its parthenogenetic
RT   daughter species: evidence from cytochrome b and ATPase 6 gene DNA
RT   sequences.";
RL   Genetica 0:0-0(1999).
CC   -!- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A
CC       DIRECT ROLE IN THE TRANSLOCATION OF PROTONS ACROSS THE MEMBRANE
CC       (BY SIMILARITY).
CC   -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC       CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC       SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC       HAS THREE MAIN SUBUNITS: A, B AND C (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
DR   EMBL; AF147804; AAF73120.1; -.
DR   InterPro; IPR000568; ATPsynt_Asub.
DR   Pfam; PF00119; ATP-synt_A; 1.
DR   PRINTS; PR00123; ATPASEA.
DR   TIGRFAMs; TIGR01131; ATP_synt_6_or_A; 1.
DR   PROSITE; PS00449; ATPASE_A; 1.
KW   CF(0); Hydrogen ion transport; Mitochondrion; Transmembrane.
FT   NON_TER       1      1
FT   NON_TER     198    198
SQ   SEQUENCE   198 AA;  21752 MW;  B6181D9B2032EDCF CRC64;

  Query Match             43.4%;  Score 46;  DB 8;  Length 198;
  Best Local Similarity   45.0%;  Pred. No. 8.3;
  Matches    9;  Conservative    7;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 LVRLSSCVPVALMSAMTTSS 21
           |::|:|   :|||: |||::
Db     161 LIQLTSTAVLALMNTMTTTA 180


RESULT 11
Q9MNY3
ID   Q9MNY3      PRELIMINARY;      PRT;   198 AA.
AC   Q9MNY3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
```

```
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ATP synthase A chain (EC 3.6.1.34) (Fragment).
OS   Lacerta dahli.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Scincomorpha; Lacertoidea;
OC   Lacertidae; Lacerta.
OX   NCBI_TaxID=94910;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Fu J., Murphy R.W., Darevsky I.S.;
RT   "Limited genetic variation in Lacerta mixta and its parthenogenetic
RT   daughter species: evidence from cytochrome b and ATPase 6 gene DNA
RT   sequences.";
     Genetica 0:0-0(1999).
     -!- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A
         DIRECT ROLE IN THE TRANSLOCATION OF PROTONS ACROSS THE MEMBRANE
CC       (BY SIMILARITY).
CC   -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC       CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC       SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC       HAS THREE MAIN SUBUNITS: A, B AND C (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
DR   EMBL; AF147805; AAF73121.1; -.
DR   InterPro; IPR000568; ATPsynt_Asub.
DR   Pfam; PF00119; ATP-synt_A; 1.
DR   PRINTS; PR00123; ATPASEA.
DR   TIGRFAMs; TIGR01131; ATP_synt_6_or_A; 1.
DR   PROSITE; PS00449; ATPASE_A; 1.
KW   CF(0); Hydrogen ion transport; Mitochondrion; Transmembrane.
FT   NON_TER       1      1
FT   NON_TER     198    198
SQ   SEQUENCE   198 AA;  21752 MW;  B6181D9B2032EDCF CRC64;

  Query Match             43.4%;  Score 46;  DB 8;  Length 198;
  Best Local Similarity   45.0%;  Pred. No. 8.3;
  Matches    9;  Conservative    7;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 LVRLSSCVPVALMSAMTTSS 21
           |::|:|   :|||: |||::
Db     161 LIQLTSTAVLALMNTMTTTA 180


RESULT 12
Q9MM88
ID   Q9MM88      PRELIMINARY;      PRT;   198 AA.
AC   Q9MM88;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ATP synthase A chain (EC 3.6.1.34) (Fragment).
OS   Darevskia derjugini.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Scincomorpha; Lacertoidea;
OC   Lacertidae; Darevskia.
OX   NCBI_TaxID=122334;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Fu J.;
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A
CC       DIRECT ROLE IN THE TRANSLOCATION OF PROTONS ACROSS THE MEMBRANE
CC       (BY SIMILARITY).
CC   -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC       CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC       SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC       HAS THREE MAIN SUBUNITS: A, B AND C (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
DR   EMBL; AF206165; AAF70411.1; -.
DR   InterPro; IPR000568; ATPsynt_Asub.
DR   Pfam; PF00119; ATP-synt_A; 1.
DR   PRINTS; PR00123; ATPASEA.
DR   TIGRFAMs; TIGR01131; ATP_synt_6_or_A; 1.
DR   PROSITE; PS00449; ATPASE_A; 1.
KW   CF(0); Hydrogen ion transport; Mitochondrion; Transmembrane.
FT   NON_TER       1      1
FT   NON_TER     198    198
SQ   SEQUENCE   198 AA;  21897 MW;  2E614F5333B97C1F CRC64;

  Query Match             43.4%;  Score 46;  DB 8;  Length 198;
  Best Local Similarity   45.0%;  Pred. No. 8.3;
  Matches    9;  Conservative    7;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 LVRLSSCVPVALMSAMTTSS 21
           |::|:|   :|||: |||::
Db     161 LIQLTSTAVLALMNTMTTTA 180


RESULT 13
Q9MM86
ID   Q9MM86      PRELIMINARY;      PRT;   198 AA.
AC   Q9MM86;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ATP synthase A chain (EC 3.6.1.34) (Fragment).
OS   Darevskia raddei.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Scincomorpha; Lacertoidea;
OC   Lacertidae; Darevskia.
OX   NCBI_TaxID=122337;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Fu J.;
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A
CC       DIRECT ROLE IN THE TRANSLOCATION OF PROTONS ACROSS THE MEMBRANE
CC       (BY SIMILARITY).
CC   -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC       CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC       SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC       HAS THREE MAIN SUBUNITS: A, B AND C (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
DR   EMBL; AF206167; AAF70413.1; -.
DR   InterPro; IPR000568; ATPsynt_Asub.
DR   Pfam; PF00119; ATP-synt_A; 1.
DR   PRINTS; PR00123; ATPASEA.
DR   TIGRFAMs; TIGR01131; ATP_synt_6_or_A; 1.
DR   PROSITE; PS00449; ATPASE_A; 1.
KW   CF(0); Hydrogen ion transport; Mitochondrion; Transmembrane.
FT   NON_TER       1      1
FT   NON_TER     198    198
SQ   SEQUENCE   198 AA;  21826 MW;  50425BC588105D12 CRC64;

  Query Match             43.4%;  Score 46;  DB 8;  Length 198;
  Best Local Similarity   45.0%;  Pred. No. 8.3;
  Matches    9;  Conservative    7;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 LVRLSSCVPVALMSAMTTSS 21
           |::|:|   :|||: |||::
Db     161 LIQLTSTAVLALMNTMTTTA 180


RESULT 14
Q9MM85
ID   Q9MM85      PRELIMINARY;      PRT;   198 AA.
AC   Q9MM85;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
```

```
DE   ATP synthase A chain (EC 3.6.1.34) (Fragment).
OS   Darevskia brauneri.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Scincomorpha; Lacertoidea;
OC   Lacertidae; Darevskia.
OX   NCBI_TaxID=122332;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Fu J.;
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A
CC       DIRECT ROLE IN THE TRANSLOCATION OF PROTONS ACROSS THE MEMBRANE
CC       (BY SIMILARITY).
CC   -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC       CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC       SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC       HAS THREE MAIN SUBUNITS: A, B AND C (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
DR   EMBL; AF206168; AAF70414.1; -.
DR   InterPro; IPR000568; ATPsynt_Asub.
DR   Pfam; PF00119; ATP-synt_A; 1.
DR   PRINTS; PR00123; ATPASEA.
DR   TIGRFAMs; TIGR01131; ATP_synt_6_or_A; 1.
DR   PROSITE; PS00449; ATPASE_A; 1.
KW   CF(0); Hydrogen ion transport; Mitochondrion; Transmembrane.
FT   NON_TER       1      1
FT   NON_TER     198    198
SQ   SEQUENCE   198 AA;  21708 MW;  895DA993435DA0AF CRC64;

  Query Match             43.4%;  Score 46;  DB 8;  Length 198;
  Best Local Similarity   45.0%;  Pred. No. 8.3;
  Matches    9;  Conservative    7;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 LVRLSSCVPVALMSAMTTSS 21
           |::|:|   :|||: |||::
Db     161 LIQLTSTAVLALMNTMTTTA 180


RESULT 15
Q8XHZ2
ID   Q8XHZ2      PRELIMINARY;      PRT;   230 AA.
AC   Q8XHZ2;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Two-component response regulator.
GN   CPE2332.
OS   Clostridium perfringens.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC   Clostridiales; Clostridiaceae; Clostridium.
     NCBI_TaxID=1502;
     [1]
     SEQUENCE FROM N.A.
RC   STRAIN=13 / TYPE A;
RX   PubMed=11792842;
RA   Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA   Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT   "Complete genome sequence of Clostridium perfringens, an anaerobic
RT   flesh-eater.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR   EMBL; AP003193; BAB82038.1; -.
DR   InterPro; IPR001789; Response_reg.
DR   InterPro; IPR001867; Trans_reg_C.
DR   Pfam; PF00072; response_reg; 1.
DR   Pfam; PF00486; trans_reg_C; 1.
DR   ProDom; PD000039; Response_reg; 1.
DR   ProDom; PD000329; Trans_reg_C; 1.
DR   SMART; SM00448; REC; 1.
DR   PROSITE; PS50110; RESPONSE_REGULATORY; 1.
KW   Complete proteome.
SQ   SEQUENCE   230 AA;  26141 MW;  5D8827B2CFC56E2A CRC64;

  Query Match             43.4%;  Score 46;  DB 16;  Length 230;
  Best Local Similarity   41.2%;  Pred. No. 9.6;
  Matches    7;  Conservative    6;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 LVRLSSCVPVALMSAMT 18
           ::|  ||||: :::| |
Db      65 VIRAKSCVPIIMITAKT 81


Search completed: May  7, 2003, 09:30:57
Job time : 31 secs
```

GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: May 7, 2003, 09:29:25 ; Search time 17 Seconds
(without alignments)
124.505 Million cell updates/sec

Title: US-09-674-973A-17
Perfect score: 106
Sequence: 1 SLVRLSSCVPVALMSAMTTSSSQ 23

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 349150 seqs, 92025710 residues

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	106	100.0	34	10	US-09-878-905-3	Sequence 3, Appli
2	44	41.5	34	10	US-09-864-761-40969	Sequence 40969, A
3	43	40.6	43	9	US-09-948-820-71	Sequence 71, Appl
4	43	40.6	323	10	US-09-943-002-12	Sequence 12, Appl
5	41.5	39.2	459	9	US-10-102-806-469	Sequence 469, App
6	41	38.7	137	10	US-09-765-272-112	Sequence 112, App
7	41	38.7	1203	9	US-10-067-457-3	Sequence 3, Appli
8	40	37.7	76	9	US-10-091-504-947	Sequence 947, App
9	40	37.7	76	10	US-09-764-869-947	Sequence 947, App
10	40	37.7	179	9	US-09-999-686-37	Sequence 37, Appl
11	40	37.7	190	10	US-09-925-302-487	Sequence 487, App
12	40	37.7	192	10	US-09-864-761-34312	Sequence 34312, A
13	40	37.7	261	9	US-09-999-686-36	Sequence 36, Appl
14	40	37.7	271	9	US-09-999-686-21	Sequence 21, Appl
15	40	37.7	301	9	US-09-999-686-35	Sequence 35, Appl
16	40	37.7	383	9	US-09-999-686-34	Sequence 34, Appl
17	40	37.7	399	9	US-09-738-626-6776	Sequence 6776, Ap
18	40	37.7	412	9	US-09-999-686-33	Sequence 33, Appl
19	40	37.7	429	9	US-10-006-915-1	Sequence 1, Appli
20	40	37.7	494	9	US-09-999-686-32	Sequence 32, Appl
21	40	37.7	543	9	US-09-999-686-2	Sequence 2, Appli
22	40	37.7	543	9	US-09-999-686-38	Sequence 38, Appl
23	40	37.7	543	9	US-09-999-686-39	Sequence 39, Appl
24	40	37.7	543	10	US-09-919-497-58	Sequence 58, Appl
25	40	37.7	549	9	US-09-999-686-31	Sequence 31, Appl
26	40	37.7	580	10	US-09-808-387-36	Sequence 36, Appl
27	40	37.7	1367	10	US-09-801-368-108	Sequence 108, App
28	39	36.8	226	9	US-09-738-626-4340	Sequence 4340, Ap
29	39	36.8	226	10	US-09-990-337-2	Sequence 2, Appli
30	39	36.8	308	10	US-09-036-613-5	Sequence 5, Appli
31	39	36.8	483	9	US-10-050-704-272	Sequence 272, App
32	39	36.8	517	9	US-09-738-626-4534	Sequence 4534, Ap
33	39	36.8	517	9	US-09-738-626-6001	Sequence 6001, Ap
34	39	36.8	528	9	US-09-738-626-6182	Sequence 6182, Ap
35	39	36.8	650	9	US-09-951-061A-94	Sequence 94, Appl
36	39	36.8	662	9	US-09-951-061A-92	Sequence 92, Appl
37	39	36.8	766	10	US-09-925-301-1276	Sequence 1276, Ap
38	38.5	36.3	176	12	US-10-078-929-86	Sequence 86, Appl
39	38	35.8	31	10	US-09-864-761-39051	Sequence 39051, A
40	38	35.8	34	10	US-09-864-761-40561	Sequence 40561, A
41	38	35.8	88	9	US-10-046-938-24	Sequence 24, Appl
42	38	35.8	169	9	US-10-172-399-10	Sequence 10, Appl
43	38	35.8	216	10	US-09-747-155-69	Sequence 69, Appl
44	38	35.8	216	10	US-09-747-155-106	Sequence 106, App
45	38	35.8	311	10	US-09-886-055-37	Sequence 37, Appl

ALIGNMENTS

RESULT 1
US-09-878-905-3
; Sequence 3, Application US/09878905
; Patent No. US20020064786A1
; GENERAL INFORMATION:
; APPLICANT: Markowitz, Sanford D
; APPLICANT: Brattain, Michael G
; APPLICANT: Willson, James K.V.
; TITLE OF INVENTION: CANCER DIAGNOSIS, PROGNOSIS AND THERAPY BASED ON
; TITLE OF INVENTION: MUTATION OF RECEPTOR
; FILE REFERENCE: 062361.0108
; CURRENT APPLICATION NUMBER: US/09/878,905
; CURRENT FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: 08/417,867
; PRIOR FILING DATE: 1995-04-07
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 34
; TYPE: PRT
; ORGANISM: human
US-09-878-905-3

Query Match 100.0%; Score 106; DB 10; Length 34;
Best Local Similarity 100.0%; Pred. No. 6e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLVRLSSCVPVALMSAMTTSSSQ 23
||||||||||||||||||||||||
Db 1 SLVRLSSCVPVALMSAMTTSSSQ 23

RESULT 2
US-09-864-761-40969
; Sequence 40969, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO

```
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 40969
;   LENGTH: 34
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL161781.2
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.5
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =  16
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 8.8
    OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL =  15
    OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL =  11
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.8
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =  13
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =   5
;   OTHER INFORMATION: EST_HUMAN HIT: AA428342.1, EVALUE 4.00e-06
US-09-864-761-40969

  Query Match             41.5%;  Score 44;  DB 10;  Length 34;
  Best Local Similarity   50.0%;  Pred. No. 2.9;
  Matches    9;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       4 RLSSCVPVALMSAMTTSS 21
           | |:| |||  |: :|:|
Db      11 RQSTCCPVAYRSSWSTTS 28


RESULT 3
US-09-948-820-71
; Sequence 71, Application US/09948820
; Publication No. US20030050460A1
; GENERAL INFORMATION:
;  APPLICANT: Ni et al.
;  TITLE OF INVENTION: 31 Human Secreted Proteins
;  FILE REFERENCE: PZ034P1
;  CURRENT APPLICATION NUMBER: US/09/948,820
;  CURRENT FILING DATE:  2001-09-10
;  PRIOR APPLICATION NUMBER: US/09/565,391
;  PRIOR FILING DATE: 2000-05-05
;  PRIOR APPLICATION NUMBER: PCT/US99/26409
;  PRIOR FILING DATE: 1999-11-09
;  PRIOR APPLICATION NUMBER: 60/108,207
;  PRIOR FILING DATE: 1998-11-12
;  NUMBER OF SEQ ID NOS: 115
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 71
;   LENGTH: 43
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (43)
;   OTHER INFORMATION: Xaa equals stop translation
US-09-948-820-71

  Query Match             40.6%;  Score 43;  DB 9;  Length 43;
  Best Local Similarity   52.9%;  Pred. No. 5.3;
  Matches    9;  Conservative    4;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       7 SCVPVALMSAMTTSSSQ 23
           :|| :  |: ||||||:
Db      20 TCVLLLSMTGMTTSSSR 36


RESULT 4
US-09-943-002-12
; Sequence 12, Application US/09943002
; Patent No. US20020045734A1
; GENERAL INFORMATION:
;  APPLICANT: Duncan, Roy
;  TITLE OF INVENTION: NOVEL REOVIRUS-DERIVED PROTEINS AND USES THEREFOR
;  FILE REFERENCE: 78973-1C
;  CURRENT APPLICATION NUMBER: US/09/943,002
;  CURRENT FILING DATE:  2001-08-31
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 12
;   LENGTH: 323
;   TYPE: PRT
;   ORGANISM: Nelson Bay virus
US-09-943-002-12

  Query Match             40.6%;  Score 43;  DB 10;  Length 323;
  Best Local Similarity   50.0%;  Pred. No. 46;
  Matches   10;  Conservative    3;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 SLVRLSSCVPVALMSAMTTS 20
           ||  ||| :| :| | :| |
Db     149 SLNSLSSSIPTSLASPLTVS 168


RESULT 5
US-10-102-806-469
; Sequence 469, Application US/10102806
; Publication No. US20030054421A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA103P1C1
;  CURRENT APPLICATION NUMBER: US/10/102,806
;  CURRENT FILING DATE:  2002-03-22
;  PRIOR APPLICATION NUMBER: 09/925,298
;  PRIOR FILING DATE: 2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05881
;  PRIOR FILING DATE: 2000-03-08
```

```
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 846
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 469
;   LENGTH: 459
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-102-806-469

  Query Match            39.2%;  Score 41.5;  DB 9;  Length 459;
  Best Local Similarity  58.3%;  Pred. No. 1.2e+02;
  Matches   14;  Conservative    0;  Mismatches    7;  Indels    3;  Gaps    1;

Qy       3 VRLSSCVPVALMSA---MTTSSSQ 23
           |||||  || |  |   ||||  |
         11 VRLSSPSPVCLPPAAATMTTSIRQ 34


RESULT 6
US-09-765-272-112
; Sequence 112, Application US/09765272
; Patent No. US20020061545A1
;   GENERAL INFORMATION:
;        APPLICANT: Choi et. al.
;        TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
;        NUMBER OF SEQUENCES: 452
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Avenue
;             CITY: Rockville
;             STATE: Maryland
;             COUNTRY: USA
;             ZIP: 20850
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;             COMPUTER: HP Vectra 486/33
;             OPERATING SYSTEM: MSDOS version 6.2
;             SOFTWARE: ASCII Text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/765,272
;             FILING DATE: 22-Jan-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/961,083
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Brookes, A. Anders
;             REGISTRATION NUMBER: 36,373
;             REFERENCE/DOCKET NUMBER: PB340P2
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (301) 309-8504
;             TELEFAX: (301) 309-8512
;   INFORMATION FOR SEQ ID NO: 112:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 137 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 112:
US-09-765-272-112

  Query Match            38.7%;  Score 41;  DB 10;  Length 137;
  Best Local Similarity  45.0%;  Pred. No. 38;
  Matches    9;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       2 LVRLSSCVPVALMSAMTTSS 21
           |:|| :|| ::|     |||
Db      34 LLRLENCVAISLGKICWTSS 53


RESULT 7
US-10-067-457-3
; Sequence 3, Application US/10067457
; Publication No. US20030082513A1
; GENERAL INFORMATION:
;  APPLICANT: Aventis Pharma Deutschland GmbH
;  TITLE OF INVENTION: Process for identifying substances which modulate the
;  TITLE OF INVENTION:  activity of hyperpolarization-activated cation channels
;  FILE REFERENCE: AVE D-2000/A006
;  CURRENT APPLICATION NUMBER: US/10/067,457
;  CURRENT FILING DATE:  2002-04-09
;  PRIOR APPLICATION NUMBER: US/09/779,587
;  PRIOR FILING DATE: 2001-02-09
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 1203
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-067-457-3

  Query Match            38.7%;  Score 41;  DB 9;  Length 1203;
  Best Local Similarity  55.0%;  Pred. No. 3.9e+02;
  Matches   11;  Conservative    1;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       2 LVRLSSCVPVALMSAMTTSS 21
           | || | :| || ||   ||
Db     826 LKRLQSLIPSALGSASPASS 845


RESULT 8
US-10-091-504-947
; Sequence 947, Application US/10091504
; Publication No. US20030059908A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC007C1
;  CURRENT APPLICATION NUMBER: US/10/091,504
;  CURRENT FILING DATE:  2002-03-07
;  NUMBER OF SEQ ID NOS: 2442
;  Prior Application removed - See File Wrapper or Palm
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 947
;   LENGTH: 76
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-091-504-947

  Query Match            37.7%;  Score 40;  DB 9;  Length 76;
  Best Local Similarity  42.9%;  Pred. No. 29;
  Matches    9;  Conservative    5;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       3 VRLSSCVPVALMSAMTTSSSQ 23
           || | |  |  :|::||:| :
Db      12 VRNSMCRSVESVSSLTTASQE 32


RESULT 9
US-09-764-869-947
; Sequence 947, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC007
;  CURRENT APPLICATION NUMBER: US/09/764,869
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - refer to PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 2442
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 947
;   LENGTH: 76
;   TYPE: PRT
```

```
;   ORGANISM: Homo sapiens
US-09-764-869-947

  Query Match             37.7%;  Score 40;  DB 10;  Length 76;
  Best Local Similarity   42.9%;  Pred. No. 29;
  Matches    9;  Conservative    5;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       3 VRLSSCVPVALMSAMTTSSSQ 23
           || | |  |  :|:::||:| :
Db      12 VRNSMCRSVESVSSLTTASQE 32


RESULT 10
US-09-999-686-37
; Sequence 37, Application US/09999686
; Publication No. US20030028000A1
; GENERAL INFORMATION:
;  APPLICANT: Aziz, Nazneen
;  APPLICANT:  Hedley, Mary Lynne
;  APPLICANT:  Urban, Robert G.
;  APPLICANT:  Tomlinson, Andrew J.
;  APPLICANT:  Cole, Geoffrey
;  TITLE OF INVENTION: CYP1B1 NUCLEIC ACIDS AND METHODS OF USE
;  FILE REFERENCE: 08191-021001
;  CURRENT APPLICATION NUMBER: US/09/999,686
;  CURRENT FILING DATE:  2001-10-31
;  PRIOR APPLICATION NUMBER: 60/298,428
;  PRIOR FILING DATE: 2001-06-15
;  PRIOR APPLICATION NUMBER: 60/261,719
;  PRIOR FILING DATE: 2001-01-12
;  PRIOR APPLICATION NUMBER: 60/244,501
;  PRIOR FILING DATE: 2000-10-31
;  NUMBER OF SEQ ID NOS: 56
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 37
;   LENGTH: 179
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-999-686-37

  Query Match             37.7%;  Score 40;  DB 9;  Length 179;
  Best Local Similarity   45.0%;  Pred. No. 72;
  Matches    9;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       3 VRLSSCVPVALMSAMTTSSS 22
           :| || ||| :  | | ::|
Db      97 MRFSSFVPVTIPHATTANTS 116


RESULT 11
US-09-925-302-487
; Sequence 487, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA104
;  CURRENT APPLICATION NUMBER: US/09/925,302
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05918
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 896
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 487
;   LENGTH: 190
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (106)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-925-302-487

  Query Match             37.7%;  Score 40;  DB 10;  Length 190;
  Best Local Similarity   45.0%;  Pred. No. 77;
  Matches    9;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       3 VRLSSCVPVALMSAMTTSSS 22
           :| || ||| :  | | ::|
Db      63 MRFSSFVPVTIPHATTANTS 82


RESULT 12
US-09-864-761-34312
; Sequence 34312, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34312
;   LENGTH: 192
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC009229.1
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.3
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.5
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =   2
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.9
```

; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.6
; OTHER INFORMATION: EST_HUMAN HIT: AU120297.1, EVALUE 1.00e-102
; OTHER INFORMATION: SWISSPROT HIT: Q16678, EVALUE 1.00e-111
US-09-864-761-34312

Query Match 37.7%; Score 40; DB 10; Length 192;
Best Local Similarity 45.0%; Pred. No. 78;
Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy 3 VRLSSCVPVALMSAMTTSSS 22
:| || ||| : | | ::|
41 MRFSSFVPVTIPHATTANTS 60

RESULT 13
US-09-999-686-36
; Sequence 36, Application US/09999686
; Publication No. US20030028000A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Nazneen
; APPLICANT: Hedley, Mary Lynne
; APPLICANT: Urban, Robert G.
; APPLICANT: Tomlinson, Andrew J.
; APPLICANT: Cole, Geoffrey
; TITLE OF INVENTION: CYP1B1 NUCLEIC ACIDS AND METHODS OF USE
; FILE REFERENCE: 08191-021001
; CURRENT APPLICATION NUMBER: US/09/999,686
; CURRENT FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: 60/298,428
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/261,719
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: 60/244,501
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-999-686-36

Query Match 37.7%; Score 40; DB 9; Length 261;
Best Local Similarity 45.0%; Pred. No. 1.1e+02;
Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy 3 VRLSSCVPVALMSAMTTSSS 22
:| || ||| : | | ::|
Db 97 MRFSSFVPVTIPHATTANTS 116

RESULT 14
US-09-999-686-21
; Sequence 21, Application US/09999686
; Publication No. US20030028000A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Nazneen
; APPLICANT: Hedley, Mary Lynne
; APPLICANT: Urban, Robert G.
; APPLICANT: Tomlinson, Andrew J.
; APPLICANT: Cole, Geoffrey
; TITLE OF INVENTION: CYP1B1 NUCLEIC ACIDS AND METHODS OF USE
; FILE REFERENCE: 08191-021001
; CURRENT APPLICATION NUMBER: US/09/999,686
; CURRENT FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: 60/298,428
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/261,719
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: 60/244,501
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 271
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-999-686-21

Query Match 37.7%; Score 40; DB 9; Length 271;
Best Local Similarity 45.0%; Pred. No. 1.1e+02;
Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy 3 VRLSSCVPVALMSAMTTSSS 22
:| || ||| : | | ::|
Db 117 MRFSSFVPVTIPHATTANTS 136

RESULT 15
US-09-999-686-35
; Sequence 35, Application US/09999686
; Publication No. US20030028000A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Nazneen
; APPLICANT: Hedley, Mary Lynne
; APPLICANT: Urban, Robert G.
; APPLICANT: Tomlinson, Andrew J.
; APPLICANT: Cole, Geoffrey
; TITLE OF INVENTION: CYP1B1 NUCLEIC ACIDS AND METHODS OF USE
; FILE REFERENCE: 08191-021001
; CURRENT APPLICATION NUMBER: US/09/999,686
; CURRENT FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: 60/298,428
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/261,719
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: 60/244,501
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 301
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-999-686-35

Query Match 37.7%; Score 40; DB 9; Length 301;
Best Local Similarity 45.0%; Pred. No. 1.3e+02;
Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy 3 VRLSSCVPVALMSAMTTSSS 22
:| || ||| : | | ::|
Db 219 MRFSSFVPVTIPHATTANTS 238

Search completed: May 7, 2003, 09:31:42
Job time : 18 secs